University of Bath

PHD

The coalescent structure of continuous-time Galton-Watson trees

Johnston, Samuel

Award date:
2018

Awarding institution:
University of Bath

The coalescent structure of continuous-time Galton-Watson trees

submitted by

S. G. G. Johnston

for the degree of Doctor of Philosophy

of the

University of Bath

Department of Mathematical Sciences

September 2017

Contents

SUMMARY

Suppose we have a continuous-time Galton-Watson tree run until time T. Conditioned on the event that there are at least k individuals alive at time T, pick uniformly k distinct particles. The goal of this thesis is to characterise the ancestry of these k individuals, with a particular emphasis on the asymptotics as $T \to \infty$.

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Matt Roberts and Simon Harris, for their continuous support, enthusiasm and encouragement. This thesis wouldn't exist without their vast knowledge and guidance, and most of all, their incredible ideas.

I'm grateful to my examiners Andreas Kyprianou and in particular Amaury Lambert, who pointed out several helpful references and demystified several formula.

Thanks to my colleagues and friends, in particular Marcus Kaiser for his technical help and Francis Lane for steering me towards better mathematical taste.

I am also in debt to Tom Sanders and David Steinsaltz - without their inspiration I wouldn't have chosen to pursue mathematics.

Finally, thanks go to my parents for their continual love and support.

Chapter 1

Introduction

Suppose we have a continuous-time Galton-Watson tree. Conditioned on the event that there are at least k individuals alive at time T, pick uniformly k distinct particles from those alive at time T. The goal of this thesis is to characterise the ancestry of these k individuals, with a particular emphasis on the asymptotics as $T \to \infty$.

1.1. Definitions

Here we give a formal definition of the continuous-time Galton-Watson tree, lifted from [20]. Let L be a random variable taking values in $\{0, 1, 2, ...\}$ and let $f(s) = \mathbb{E}[s^L]$ be its generating function. Under a probability measure $\mathbb{P}$, we begin with one particle, the root, which we label $\emptyset$. This particle waits an exponential amount of time $\tau_\emptyset$ with parameter 1, and then instantaneously dies and gives birth to some offspring with labels $1, 2, \ldots, L_\emptyset$, where $L_\emptyset$ is an independent copy of the random variable L. To be precise, at the time $\tau_\emptyset$ the particle $\emptyset$ is no longer alive and its offspring are. These offspring then repeat, independently, this behaviour: each particle u waits an independent exponential amount of time with parameter 1 before dying and giving birth to offspring $u1, u2, \ldots, uL_u$ where L_u is an independent copy of L, and so on. This construction gives rise to a random tree which we call the continuous-time Galton-Watson tree with offspring distribution L, or L-tree for short.

Denote by $\mathcal{N}_t$ the set of all particles alive at time t, let $N_t = \#\mathcal{N}_t$ be the number alive and let $F_t(s) = \mathbb{E}[s^{N_t}]$ be the generating function of the process. For a particle $u \in \mathcal{N}_T$ we let τ_u be the time of its death, and define $\tau_u(T) = \tau_u \wedge T$. If u is an ancestor of v, we write $u \leq v$, and if u is a *strict* ancestor of v (i.e. $u \leq v$ and $u \neq v$) then we write $u < v$.

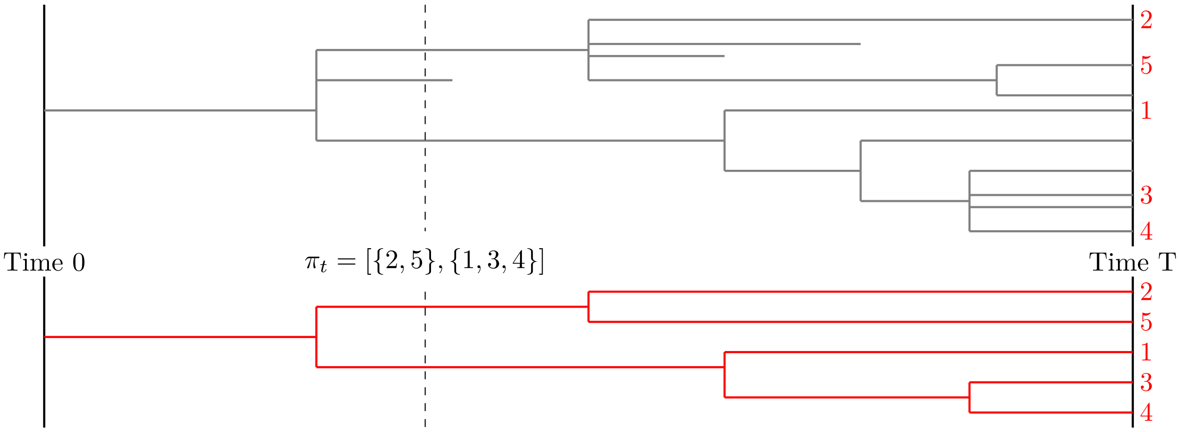
For a particle $u \in \mathcal{N}_T$ and $t \leq T$, let $u(t)$ be the unique ancestor of u that was alive at time t. For two particles $u \neq v \in \mathcal{N}_T$, let $\sigma(u, v)$ be the last time at which they shared a common ancestor,

$$\sigma(u, v) = \sup\{t \geq 0 : u(t) = v(t)\}.$$

For $T > 0$ we condition on the event $\{N_T \geq k\}$, and pick k particles alive at time T, uniformly and without replacement, labelling them $1, 2, ..., k$. For $t \in [0, T]$, let $i(t)$ be the unique time-t ancestor of the particle labelled i. For each time $t \in [0, T]$, we define the equivalence relation

$$i \sim_{\pi_t} j \iff i(t) = j(t)$$

We let $\pi_t^{k,L,T}$ denote the random partition of $\{1, ..., k\}$ corresponding to this equivalence relation. Clearly $(\pi_t) := (\pi_t^{k,L,T})_{t\in[0,T]}$ is a right-continuous partition-valued process characterising the entire ancestral tree of the k particles labelled at time T.



In certain cases it is more natural to consider time going backwards from time T. With this motivation, define $(\rho_t) = (\rho_t^{k,L,T})_{t\in[0,T]}$ by setting $\hat{\rho}_t = \pi_{T-t}$, and letting (ρ_t) be the right-continuous modification of $(\hat{\rho}_t)$. The process (ρ_t) is a *coalescent process* in the sense that blocks merge as time passes.

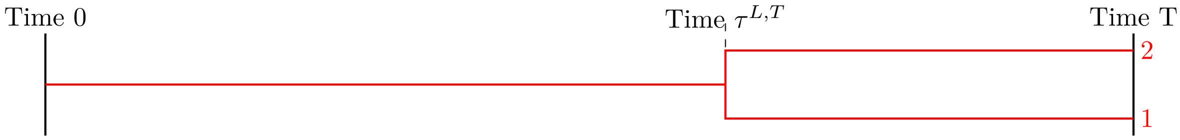
1.2. Summary of Results

We will be particularly interested in the behaviour of the process $(\pi_t^{k,L,T})_{t\in[0,T]}$ as $T \to \infty$, finding that there are marked differences in the qualititative behaviour depending on the mean number of offspring $m = \mathbb{E}[L]$. Before we look at our results in full generality, we begin by looking at the case $k = 2$, where the majority of literature on coalescence in Galton-Watson trees lies. We reproduce in more detail the introductory discussion on the case $k = 2$ from [24].

For the sake of fluency, sometimes we state in continuous-time results which were actually proven by their authors in discrete-time. In our discussion of embeddability in section 1.3, we give a justification for this conversion.

1.2.1 An example: the time to most recent common ancestor of $k = 2$ individuals chosen uniformly at time T

The case $k = 2$ amounts to choosing two individuals from a tree with offspring distribution L at a time T, and studying the time $\tau^{L,T} \in [0, T]$ at which they last shared a common ancestor. Equivalently, $\tau^{L,T}$ is the time that the single block $[\{1, 2\}]$ of the process $(\pi_t^{2,L,T})_{t\in[0,T]}$ splits into the pair of singletons $[\{1\}, \{2\}]$.



In this direction, the following result (which we will generalise significantly later) was proved by Lambert [31] in discrete-time. Le gave the continuous-time analogue in [36].

Lemma 1.2.1 (Lambert [31], Corollary 1)**.**

$$\mathbb{P}\Big(\tau^{L,T} \in [t,T]\Big) = \frac{1}{\mathbb{P}(N_T \geq 2)} \int_0^1 (1-s) \frac{F''_{T-t}(s)}{F'_{T-t}(s)} F'_T(s) ds \tag{1.1}$$

where we recall $F_t(s) = \mathbb{E}[s^{N_t}]$ is the generating function for the number alive in the process.

Although this result gives a powerful implicit characterisation of the distribution of $\tau^{L,T}$, it is difficult to infer qualitative properties of this random variable directly from the formula above. By sending $T \to \infty$, however, it is possible to gain a more intuitive insight. Unsurprisingly, different qualitative behaviours arise depending on whether the underlying Galton-Watson process is supercritical, critical, or subcritical. Letting $m := \mathbb{E}[L]$ denote the mean number of offspring, these cases correspond to $m > 1$, $m = 1$, and $m < 1$ respectively.

Buhler first observed in [11], that when two individuals are picked from a supercritical tree at a large time, their most recent common ancestor was a member of one of the generations near the start. Indeed, Athreya [4] showed (albeit in discrete-time) that when $1 < \mathbb{E}[L] < \infty$, the time $\tau^{L,T}$ remains near the beginning of the interval $[0,T]$ even when T is very large, and consequently we have the distributional convergence

$$\tau^{L,T} \xrightarrow{D} \bar{\tau}^L \in [0,\infty).$$

Under an $L \log_+ L$ condition, the special case $k = 2$ of the supercritical theorem in [24], proven in chapter 2, allows us to provide a novel characterisation of the law of the limit variable $\bar{\tau}^L$.

Theorem 1.2.2 (Johnston [24], Theorem 2.4, $k = 2$)**.** Suppose $\mathbb{E}[L] > 1$ and the Kesten-Stigum condition $\mathbb{E}[L \log_+ L] < \infty$ holds. Then

$$\tau^{L,T} \xrightarrow{D} \bar{\tau}^L \in [0,\infty),$$

and the limit variable $\bar{\tau}^L$ has probability density given by the integral representation

$$\mathbb{P}(\bar{\tau}^L \in dt) = e^{-(m-1)t} \int_0^\infty \frac{v\varphi^1(v)}{1-\varphi^0(\infty)} \varphi^1(ve^{-(m-1)t}) f''\Big(\varphi^0(ve^{-(m-1)t})\Big) dv, \ t \in [0,\infty) \tag{1.2}$$

where $\varphi^j(v) = \mathbb{E}[W^j e^{-vW}]$, $W = \lim_{t\to\infty} N_t e^{-(m-1)t}$ is the martingale limit, and $(1 - \varphi^0(\infty))$ is the survival probability.

By way of example, an explicit formula is available when $L = 2$ almost surely, revealing surprisingly detailed structure.

Example 1.2.1. We call our process a standard Yule tree if $f(s) = s^2$. In this case, it is well known that the martingale limit W is a unit mean exponential, and hence

$$f''(u) \equiv 2,\ \varphi^0(v) = \frac{1}{(1+v)},\ \varphi^1(v) = \frac{1}{(1+v)^2}. \tag{1.3}$$

Inserting this into (1.2), a calculation shows that the limit variable $\bar{\tau}^{\mathsf{Yule}}$ has an interesting and nontrivial probability density function

$$\mathbb{P}(\bar{\tau}^{\mathsf{Yule}} \in dt) = \frac{2e^t}{(e^t-1)^3}\Big[(t-2)e^t + (t+2)\Big]dt,\ t \in [0,\infty). \tag{1.4}$$

The subcritical case is completely different. On the rare event that a subcritical tree manages to survive until a large time T, the number of individuals alive converges to a quasi-stationary limit (see [5, Chapter I, Section 8]). Furthermore, every individual alive is descended from a single ancestral lineage that survived for the majority of $[0,T]$. Where in the supercritical case, the common ancestor of two individuals chosen at a large time T last existed near the beginning of time, in the subcritical case the common ancestor last existed very near T.

Indeed, Lambert showed in [31] that the difference $\upsilon^{L,T} := T - \tau^{L,T}$ satisfies the distributional convergence

$$\upsilon^{L,T} \xrightarrow{D} \bar{\upsilon}^L \in [0,\infty). \tag{1.5}$$

Lambert also gave an implicit formula for the distribution of the limit variable $\bar{\upsilon}^L$ (see equation (1.34) below), which Le [36] inverted to give the formula

$$\mathbb{P}(\bar{\upsilon}^L \in dt) = \int_0^1 \frac{(1-s)}{\mathbb{P}(W \geq 2)} B'(s)F_t'(s)f''(F_t(s))ds,\ t \in [0,\infty), \tag{1.6}$$

where W is the quasi-stationary limit variable:

$$\mathbb{P}(W = n) = \lim_{t\to\infty} \mathbb{P}(N_t = n | N_t \geq 1),$$

and $B(s) = \mathbb{E}[s^W]$ is its generating function.

Moving on to the critical case, the hitherto most widely studied object in coalescence in Galton-Watson trees has been the limiting distribution of the time $\tau^{L,T}$ when $\mathbb{E}[L] = 1$. Several authors have found under a finite variance assumption that

$$\tau^{L,T}/T \to^D \bar{\tau}^{\mathsf{Crit}} \in [0,1], \tag{1.7}$$

where the variable $\bar{\tau}^{\mathsf{Crit}}$ is universal in all critical L with finite variance. A special case of Theorem 2.3 of [20] (proven in chapter 3) tells us $\bar{\tau}^{\mathsf{Crit}}$ has probability density function

$$\mathbb{P}(\bar{\tau}^{\mathsf{Crit}} \in dt) = \Big[-4\frac{\log(1-t)}{t^3} + 2\frac{\log(1-t)}{t^2} - 4\frac{1}{t^2}\Big]dt,\ t \in [0,1]. \tag{1.8}$$

Athreya [3] gave an implicit representation of (1.8) in terms of a geometric sum of exponentials, and Durrett [13] gave (1.8) as a power series. We prove the equivalence of Athreya's representation, Durrett's representation and equation (1.8) in Chapter 3. We will even see later that the universal limit variable $\bar{\tau}^{\mathsf{Crit}}$ turns up when we consider coalescence in continuous-state branching process [31].

In a generalisation of the critical setup, O'Connell [39] studied the random variable $\tau^{L,T}$ under the following near-critical scaling limit. Let $\{L_T : T \geq T_0\}$ be a collection of random variables satisfying

- $\mathbb{E}[L_T] = 1 + \mu/T + o(1/T),\ \mu \in \mathbb{R}$,
- $\mathbb{E}[L_T(L_T - 1)] = \sigma + o(1),\ \sigma > 0$.

Then in an L_T-tree, let $\tau^{L_T,T}$ be the distribution of the common ancestor of two uniformly chosen individuals at a time T. O'Connell related the subtree of individuals who go on to have descendents at time T with a time-changed Yule tree, ultimately proving that there exists a limit variable $\tau^{\mathsf{Crit}^\mu} \in [0,1]$ such that

$$\frac{\tau^{L_T,T}}{T} \to^D \bar{\tau}^{\mathsf{Crit}^\mu}$$

where the limit variable depends on the near-critical mean scaling μ but not the variance σ. Furthermore, $\bar{\tau}^{\mathsf{Crit}^\mu}$ has distribution function

$$\mathbb{P}(\bar{\tau}^{\mathsf{Crit}^\mu} > t) = 2\Big(\frac{e^{r\mu(1-t)} - 1}{e^{r\mu(1-t)} - e^{r\mu}}\Big) + 2\frac{(e^{r\mu} - 1)(e^{r\mu(1-t)} - 1)}{(e^{r\mu(1-s)} - e^{r\mu})^2} \log\Big(\frac{e^{r\mu} - 1}{e^{r\mu(1-t)} - 1}\Big),\ t \in [0,1]. \tag{1.9}$$

O'Connells result here corresponds to the special case $k = 2$ of our main result in [20], which gives the entire coalescent structure of these near-critical trees. Let us just remark that it is possible to recover (1.8) by taking $\mu \to 0$ in (1.9) and differentiating.

Finally, let us discuss the behaviour of $\tau^{L,T}$ when L is of infinite mean. Athreya [4] studied the offspring variables in the domain of attraction of a α-stable law which we define here informally as

$$\mathcal{M}^\alpha = \left\{L : \mathbb{E}[L] = 1,\ \mathbb{P}(L \geq n) \text{ looks like } n^{-\alpha}\right\}.$$

He goes on to prove that infinite mean trees bare closer relation to subcritical than supercritical trees in the structure of their coalescence. Indeed, if $L \in \mathcal{M}^\alpha$ then

$$\upsilon^{L,T} \to^D \bar{\upsilon}^L. \tag{1.10}$$

In Chapter 4, we go on to study the coalescent structure of trees with heavy-tails, with the infinite mean offspring distribution L^* given by

$$\mathbb{P}(L^* = n) = \frac{1}{n(n-1)}$$

playing a fundamental role. Interestingly, setting $L = L^*$ furnishes the simplest fixed-T distribution of the random variable $\upsilon^{L,T}$, given by density

$$\mathbb{P}(\upsilon^{L^*,T} \in dt) = \frac{e^{-t}dt}{1 - e^{-T}}$$

Clearly $\upsilon^{L^*,T}$ converges in distribution to a $\bar{\upsilon}^{L^*}$, a unit-mean exponential. We will discuss the coalescent structure of L^*-trees again shortly.

We now move on to giving the results of this thesis which concern general k. With the exception of a special case of [24, Theorem 2.1] given in discrete-time by Grosjean and Huillet (see below), the results of sections 1.2.2 to 1.2.6 are original.

1.2.2 Formula characterising the law of $(\pi_t^{k,L,T})_{t \in [0,T]}$ for fixed-T

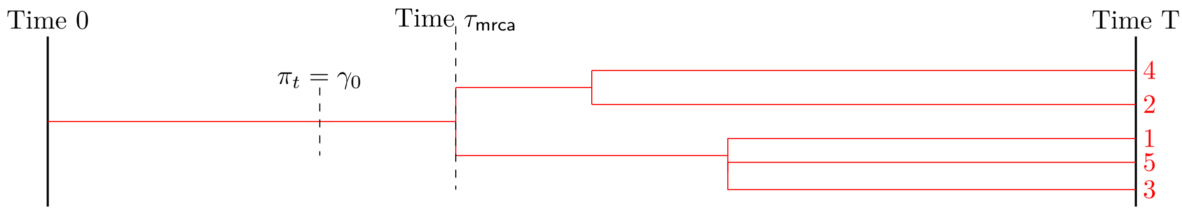
In [24], we give a complete characterisation for the law of (π_t^T) in terms of integral formula involving s-derivatives of the generating functions $f(s)$ and $F_t(s)$. For example, for a fixed $t \in [0, T]$, the following formula gives the law of the partition-valued random variable $\pi_t^{k,L,T}$.

Theorem 1.2.3 (Johnston [24], Theorem 2.1)**.** For a partition γ of $\{1, ..., k\}$ into p blocks $\Gamma_1, ..., \Gamma_p$ of sizes $k_1 + ... + k_p = k$,

$$\mathbb{P}(\pi_t^{k,L,T} = \gamma) = \int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N_T \geq k)} F_t^p(F_{T-t}(s)) \prod_{i=1}^p F_{T-t}^{k_i}(s) ds. \tag{1.11}$$

where $F_t^j(s)$ denotes the j^{th} derivative of $F_t(s)$ with respect to s.

Example 1.2.2 (The most recent common ancestor of k individuals)**.** By taking $\gamma = \gamma_0 := [\{1, ..., k\}]$ to be the partition of $\{1, ..., k\}$ into one block, we get a formula for the time to most recent common ancestor $\tau_{\mathsf{mrca}}^{k,L,T} \in [0, T]$. It should be clear from the diagram below that $\{\tau_{\mathsf{mrca}} > t\} \equiv \{\pi_t = \gamma_0\}$.



By plugging in $\gamma = \gamma_0$ into equation (1.11), and using the semigroup identity to note that

$$F_T^1(s) = F_{T-t}^1(s) F_t^1(F_{T-t}(s)),$$

we output a result by Grosjean and Huillet [17], giving the time to most recent common ancestor of k uniformly chosen individuals.

Theorem 1.2.4 (Grosjean and Huillet [17], Corollary 7)**.**

$$\mathbb{P}(\tau_{\mathsf{mrca}}^{k,L,T} > t) = \mathbb{P}(\pi_t^{k,L,T} = \gamma_0) = \int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N_T \geq k)} \frac{F_T^1(s)}{F_{T-t}^1(s)} F_{T-t}^k(s) ds. \tag{1.12}$$

The one dimensional formula (1.11) is the special case $n = 1$ of a more general formula in [24] which we state shortly, giving the finite dimensional distributions of (π_t^T):

$$\mathbb{P}(\pi_{t_1} = \gamma_1, ..., \pi_{t_n} = \gamma_n). \tag{1.13}$$

Let us just outline here some definitions involved in expressing the quantity (1.13). For two partitions α and β of $\{1, ..., k\}$, we write $\alpha \prec \beta$, and say β is a refinement of α, if every block in β is a subset of a block in α. Suppose we have a refining chain of partitions starting at one block and finishing in the singletons

$$[\{1, ..., k\}] = \gamma_0 \prec \gamma_1 \prec \gamma_2 ... \prec \gamma_n \prec \gamma_{n+1} = [\{1\}, \ldots, \{k\}],$$

then to this chain we associate a doubly-indexed set of integers $\{b_{i,j} : i = 0, 1...., n, j = 1, ..., |\gamma_i|\}$ called the breakage numbers of the sequence. The $(i, j)^{\text{th}}$ breakage number is the number of blocks the j^{th} block of the partition γ_i needs to break into to form some new blocks in γ_{i+1}. We discuss the breakage numbers in more detail in Chapter 2.

Theorem 1.2.5 (Johnston [24], Theorem 2.2)**.** Let (γ_i) be a partition chain and let $0 = t_0 < t_1 < ... < t_n < t_{n+1} = T$, and write $\Delta t_i = t_{i+1} - t_i$.

$$\mathbb{P}(\pi_{t_1} = \gamma_1, ..., \pi_{t_n} = \gamma_n) = \int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N_T \geq k)} \prod_{i=0}^{i=n} \prod_{\Gamma_{i,j} \in \gamma_i} F_{\Delta t_i}^{b_{i,j}} \Big(F_{T-t_{i+1}}(s) \Big) ds. \tag{1.14}$$

In Chapter 2, we also provide an alternate way to characterise the law of (π_t^T) is in terms of the 'split times' - the times t at which π_{t-}^T differs from π_t^T. Let $[\{1, ..., k\}] = \eta_0 \prec \eta_1 \prec ... \prec \eta_n = [\{1\}, ..., \{k\}]$ be a chain of partitions that is maximal in the sense that η_i is created from η_{i-1} by breaking precisely one block of η_{i-1} into $c_i \geq 2$ blocks in η_i. The (c_i) are called the split multiplicities, and satisfy the branch equation

$$(c_1 - 1) + ... + (c_n - 1) = k - 1.$$

For times $t_1 < ... < t_n \in [0, T]$, write

$$\{\pi : \eta_0 \prec^{dt_1} \eta_1 \prec^{dt_2} ... \prec^{dt_n} \eta_n\} = \bigcap_{i=1}^{n} \{\pi_{t_i} = \eta_{i-1}, \pi_{t_i + dt_i} = \eta_i\} \tag{1.15}$$

for the event that for each $i = 1, ..., n$, the value of (π_t^T) jumps from η_{i-1} to η_i in the time interval $(t_i, t_i + dt_i]$.

The following theorem gives the density of the event (1.15), thereby supplying an alternate characterisation of the law of (π_t^T).

Theorem 1.2.6 (Johnston [24], Theorem 2.3)**.**

$$\mathbb{P}(\pi : \eta_0 \prec^{dt_1} ... \prec^{dt_n} \eta_n) = \int_0^1 \frac{(1-s)^{k-1} F_T^1(s)}{(k-1)!\mathbb{P}(N_T \geq k)} \prod_{i=1}^{n} \Big(F_{T-t_i}^1(s)^{c_i - 1} f^{c_i}(F_{T-t_i}(s)) dt_i \Big) ds. \tag{1.16}$$

We now move on to the asymptotics of (π_t^T) as $T \to \infty$. Let $\mathcal{P}_k$ be the set of partitions on $\{1, ..., k\}$. Below, we say a collection of right-continuous $\mathcal{P}_k$-valued processes $\{\lambda^T = (\lambda_t^T)_{t\in[0,T]} : T > 0\}$ converge in distribution to a right-continuous limit process $\bar{\lambda} = (\bar{\lambda}_t)_{t\in[0,\infty)}$ as $T \to \infty$, and we write

$$(\lambda_t^T)_{t\in[0,T]} \to^D (\bar{\lambda}_t)_{t\in[0,\infty)}.$$

if for any fixed times $t_1 < .... < t_n$ and partition chain $\gamma_1 \prec ... \prec \gamma_n$ we have

$$\lim_{T\to\infty} \mathbb{P}(\lambda_{t_i}^T = \gamma_i \ \forall i) = \mathbb{P}(\bar{\lambda}_{t_i} = \gamma_i \ \forall i). \tag{1.17}$$

The limit process is constructed in such a way that (1.17) implies weak convergence in distribution under Skorokhod topology.

1.2.3 The supercritical case

In the supercritical case, it turns out that if any number $k \geq 2$ individuals are picked uniformly from those alive at a large time T, then their common ancestors last existed near the beginning of the interval $[0, T]$.

Theorem 1.2.7 (Johnston [24], Theorem 2.4)**.** Suppose $\mathbb{E}[L] > 1$ and $\mathbb{E}[L \log_+ L] < \infty$. Then as $T \to \infty$,

$$(\pi_t^{k,L,T})_{t\in[0,T]} \to^D (\bar{\pi}_t^{k,L})_{t\in[0,\infty)}$$

Furthermore, we are able to characterise the law of the limit process $(\bar{\pi}_t^{k,L})_{t\in[0,\infty)}$ through its splitting times:

$$\mathbb{P}(\bar{\pi}^{k,L} : \eta_0 \prec^{dt_1} \eta_1 \prec^{dt_2} ... \prec^{dt_n} \eta_n) = \int_0^\infty \frac{v^{k-1}\varphi^1(v)}{(k-1)!(1-\varphi^0(\infty))} \left(\prod_{i=1}^n g_{c_i}(t_i, v) dt_i \right) dv \tag{1.18}$$

where $\varphi^j(v) = \mathbb{E}[W^j e^{-vW}]$, and for each v, $g_{c_i}(\cdot, v)$ are smooth, non-negative and integrable densities in $t \in [0, \infty)$ given by

$$g_c(t, v) = \left(\varphi^1(ve^{-(m-1)t_i}) e^{-(m-1)t_i} \right)^{c-1} f^c(\varphi^0(ve^{-(m-1)t_i})).$$

We remark that case $k = 2$ recovers the formula (1.2) we saw in the introduction.

Let us take this opportunity to look at a couple of examples of coalescent structures derived from supercritical Galton-Watson trees. In the interest of generating concrete examples, the following equation by Harris [19] allows us to extract martingale transforms $\varphi(v)$ from offspring generating functions $f(s)$:

$$\bar{\varphi}(y) = (1-y) \exp\left\{ \int_1^y \frac{f'(1)-1}{f(s)-s} + \frac{1}{1-s} ds \right\}, \ y \in (0, 1]. \tag{1.19}$$

where $\bar{\varphi} : (0, 1] \to [0, \infty)$ is the inverse function of $\varphi(v)$. Harris first proved this equation in 1951 under the condition $\mathbb{E}[L^2] < \infty$, though in light of the seminal work done by Kesten and Stigum in 1966 [27], Karlin and McGregor [26] showed the martingale limit equation (1.19) holds for all supercritical L with $\mathbb{E}[L \log_+ L] < \infty$.

Example 1.2.3 (The Yule Tree)**.** The simplest possible supercritical tree is the Yule Tree, corresponding to the case $f(s) = s^2$. Here the martingale transform $\varphi(v)$ takes its most tractable form:

$$\varphi(v) = \frac{1}{1+v}. \tag{1.20}$$

Noting that

$$f^c(s) \equiv 2 \text{ for } c = 2, \; f^c(s) \equiv 0 \text{ for } c \geq 3 \tag{1.21}$$

it follows that the functions $g_c(\cdot, v)$ are non-zero only when $c = 2$ (which is unsurprising since the process $(\bar{\pi}_t^{k,\mathsf{Yule}})$ corresponds to a subtree of a binary tree). It follows we can n has to take the value $k-1$, and we can write

$$\mathbb{P}(\bar{\pi}^{k,\mathsf{Yule}} : \eta_0 \prec^{dt_1} \eta_1 \prec^{dt_2} ... \prec^{dt_{k-1}} \eta_{k-1}) = \int_0^\infty \frac{v^{k-1}}{(k-1)!(1+v)^2} \prod_{i=1}^{k-1} \frac{2e^{-t_i} dt_i}{(1+ve^{-t_i})^2} dv \tag{1.22}$$

for the splitting times.

To get an idea of what happens when the supercritical tree is non-binary, we end our discussion of the supercritical case with a glance at the next case up.

Example 1.2.4 (The continuous-time Ternary tree)**.** Suppose

$$f(s) = s^3,$$

then the martingale limit equation yields

$$\varphi(v) = \frac{1}{\sqrt{1+2v}}.$$

With $f'''(s) = 6$, and $f''(s) = 6s$, we have two non-zero split densities:

$$g_3(t|v) = \frac{6e^{-4t}}{(1+2ve^{-2t})^3}, \; g_2(t|v) = \frac{6e^{-2t}}{(1+2ve^{-2t})^2}.$$

Suppose we pick three individuals from a ternary tree at a large time, what can we say about the topology of their ancestral tree? That is, what is the probability that $(\bar{\pi}_t^{3,\mathsf{tern}})$ has a single 3-split (as opposed to two 2-splits)?

Remarkably, the probability $(\bar{\pi}_t^{3,\mathsf{tern}})$ has a 3-split at any time is

$$\begin{aligned}
\int_{t=0}^\infty \mathbb{P}\left(\pi^{3,\mathsf{tern}} : [\{1,2,3\}] \prec^{dt} [\{1\},\{2\},\{3\}]\right) &= \int_{t=0}^\infty \int_{v=0}^\infty \frac{v^2}{2(1+v)^{3/2}} \frac{6e^{-4t}}{(1+2ve^{-2t})^3} dv dt \\
&= \frac{3}{8} \frac{3\log(\frac{1}{3-2^{3/2}}) - 2^{3/2}}{\sqrt{2}} \\
&\approx 0.652
\end{aligned}$$

1.2.4 The subcritical case

We mentioned in the introduction that on the event that a subcritical tree survives until a large time T, everyone alive is descended from an ancestor who existed near to T. Indeed, when $\mathbb{E}[L] < 1$, the common ancestors of a sample taken at time T last existed near T.

Theorem 1.2.8 (Johnston [24], Theorem 2.6)**.** Suppose $\mathbb{E}[L] < 1$. Then as $T \to \infty$,

$$(\rho_t^{k,L,T})_{t\in[0,T]} \to^D (\bar{\rho}_t^{k,L})_{t\in[0,\infty)}.$$

Furthermore, we are able to characterise the law of $(\bar{\rho}_t^{k,L})_{t\in[0,\infty)}$ through its merger times. Let $[\{1\},...,\{k\}] = \eta_0 \succ \eta_1 \succ ... \succ \eta_n = [\{1,2,...,k\}]$, where η_i is obtained from η_{i-1} by combining c_i blocks of η_{i-1} to form a single block in η_i. Then for any $0 < t_1 < ... < t_n$

$$\mathbb{P}(\bar{\rho}^{k,L} : \eta_0 \succ^{dt_1} \eta_1 \succ^{dt_2} ... \succ^{dt_n} \eta_n) = \int_0^1 \frac{(1-s)^{k-1}B'(s)}{(k-1)!\mathbb{P}(W \geq k)} \prod_{i=1}^n \left\{ F'_{t_i}(s)^{c_i-1} f^{c_i}(F_{t_i}(s))dt_i \right\} ds \tag{1.23}$$

The following example concerns the most recent common ancestor of two particles in a subcritical birth-death process.

Example 1.2.5. Suppose

$$f(s) = \mathbb{E}[s^L] = a + bs^2, \text{ for some } a > b = 1-a.$$

Let $\upsilon^{\mathsf{bin}(b,a)}$ be the single merger time of the process $(\bar{\rho}_t^{2,L})$, then $\upsilon^{\mathsf{bin}(b,a)}$ has density

$$\mathbb{P}(\upsilon^{\mathsf{bin}(b,a)} \in dt) = \frac{2a\delta e^{\delta t}}{b^2(ae^{\delta t}-b)} \left\{ b(a+b-2a^{\delta t}) + a(2e^{\delta t}-1)(ae^{\delta t}-b)\log\left[1+\frac{b}{ae^{\delta t}-b}\right] \right\},$$

where $\delta = a - b$.

1.2.5 The critical case

Our main result in [20] is to show that for L in a large class of critical and near-critical variables with finite variance, the process $(\pi_t^{k,L,T})_{t\in[0,T]}$ gets stretched proportionally with the interval $[0,T]$.

Theorem 1.2.9 (Harris, Johnston and Roberts [20], Theorem 2.3 and Proposition 3.1)**.** Suppose we take the following scaling limit of the offspring random variable L:

- $\mathbb{E}[L_T] = 1 + \mu/T + o(1/T),\ \mu \in \mathbb{R}$,
- $\mathbb{E}[L_T(L_T-1)] = \sigma + o(1),\ \sigma > 0$,
- The family $\{L_T : T \geq 1\}$ is uniformly integrable.

Then

$$(\pi_{tT}^{k,L_T,T})_{t\in[0,1]} \xrightarrow{D} (\bar{\pi}_t^{\mathsf{k},\mathsf{Crit}^\mu})_{t\in[0,1]}. \tag{1.24}$$

where the limit process $(\bar{\pi}_t^{\mathsf{k},\mathsf{Crit}^\mu})$ depends on μ but not σ. In particular, the limit process $(\bar{\pi}_t^{k,\mathsf{Crit}^0})$ is a universal limit for the coalescent structure of all critical continuous-time Galton-Watson trees with finite variance.

Furthermore, the class of limit processes $\left\{\left(\bar{\pi}_t^{\mathsf{k},\mathsf{Crit}^\mu}\right) : \mu \in \mathbb{R}\right\}$ are binary trees topologically equivalent to Kingman's coalescent (first constructed in [28]).

In [20], we also provide an explicit formula for the split times of the limit processes $\left(\bar{\pi}_t^{\mathsf{k},\mathsf{Crit}^\mu}\right)_{t\in[0,1]}$. For example, in the case $\mu = 0$, the split times $0 < t_1 < ... < t_{k-1} < 1$ of $\left(\bar{\pi}_t^{\mathsf{k},\mathsf{Crit}^0}\right)_{t\in[0,1]}$ are given by the integral equation

$$f_k(t_1, ..., t_{k-1}) = k \int_0^\infty \theta^k \prod_{i=0}^{k-1} \frac{1}{(1+\theta(1-t_i))^2} d\theta \tag{1.25}$$

where $t_0 = 0$.

It turns out that the growth parameter μ only affects the law of $\left(\bar{\pi}_t^{\mathsf{k},\mathsf{Crit}^\mu}\right)$ through the following deterministic time change. Namely, the bijection on $[0, 1]$ given by

$$t \mapsto \frac{1}{\mu}\log(1 + (e^\mu - 1)t) =: a_\mu(t),$$

provides the following equivalence in law via time-change:

$$\left(\bar{\pi}_t^{k,\mathsf{Crit}^\mu}\right)_{t\in[0,1]} \equiv^{\mathsf{Law}} \left(\bar{\pi}_{a_\mu(t)}^{k,\mathsf{Crit}}\right)_{t\in[0,1]}.$$

Relationships of different trees through time-change will be a of theme of this thesis. For instance, we have the more dramatic relationship between the coalescent process of the critical tree, $\left(\bar{\pi}_t^{\mathsf{k},\mathsf{Crit}^0}\right)_{t\in[0,1]}$ and the coalescent process of the (supercritical) Yule tree $\left(\bar{\pi}_t^{\mathsf{k},\mathsf{Yule}}\right)_{t\in[0,\infty)}$, linked to one another by the deterministic time change

$$\left(\bar{\pi}_t^{k,\mathsf{Crit}^0}\right)_{t\in[0,1]} \equiv^{\mathsf{Law}} \left(\bar{\pi}_{\log\left(\frac{1}{1-t}\right)}^{k,\mathsf{Yule}}\right)_{t\in[0,1]}$$

This can be seen simply by substituting t with $\log\left(\frac{1}{1-t}\right)$ in (1.25) to recover (1.22). (The difference of a factor $\frac{k!}{2^k}$ is matter counting different tree topologies - see Amaury Lambert's fascinating discussion in [34]).

A blunt method of understanding this relationship is provided by considering renormalised population sizes. The Kolmogorov-Yaglom exponential limit law (see [5, Chapter III, Section 7]) states that for a finite variance critical process $(N_t)_{t\geq 0}$

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{N_t}{\sigma^2 t/2} > x \middle| N_t > 0\right) = e^{-x}.$$

On the other hand, by (1.20), if $(N_t^{\mathsf{Yule}})_{t\geq 0}$ is a Yule process then

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{N_t^{\mathsf{Yule}}}{e^t} > x \middle| N^{\mathsf{Yule}} > 0\right) = e^{-x}.$$

That is, in each of these processes, the renormalised sizes of populations in conditioned to survive converge to exponentials - even though they are fundamentally of different orders. It follows that in either case, selecting some individuals at a large time T and asking if they shared common ancestors at a smaller time t corresponds to a question about ratios of exponentials. In coalescent terminology, the asymptotic frequencies at fixed times of the processes $(\bar{\pi}_t^{\mathsf{Crit}^0})_{t\in[0,1]}$ and $(\bar{\pi}_t^{k,\mathsf{Yule}})_{t\in[0,\infty)}$ are ratios of exponentials - which suggests some sort of correspondence between the two objects.

A more refined argument can be given by the reduced trees argument (we call them 'purple' in [20]). This idea seems to originate in Fleischmann and Siegmund-Schultze [15] and has since propogated. To give just a brief idea here, suppose we have a critical continuous-time Galton-Watson tree conditioned to survive until a large time T. Call 'purple' any particle that goes on to have descendents alive at time T. It turns out that we can characterise the dynamics of the purple particles as an inhomogenous pure birth process, with birth rate at time tT given by

$$\frac{1}{T}\log\left(\frac{1}{1-t}\right)+o(1/T), t\in[0,1].$$

Sending $T\to\infty$, the link is suggested. Later, we appeal to results from a brilliant paper by Lageras and Sagitov examining the reduced trees of critical processes with heavy tails [30]. Let us also mention here O'Connell [38], who studies the near-critical analogue, and Yakymiv [48], who looks at the multitype case.

1.2.6 Heavy-tailed trees

For $\alpha\in(0,1]$, we consider the supercritical offspring variable $L^*_{1+\alpha}$ given by generating function

$$\mathbb{E}[s^{L^*_{1+\alpha}}]=f^*_{1+\alpha}(s)=\frac{(1-s)^{1+\alpha}-1+(1+\alpha)s}{\alpha}. \tag{1.26}$$

In the case $\alpha=0$, we just write $L^*:=L^*_1$, and

$$\mathbb{E}[s^{L_*}]=f^*(s)=s+(1-s)\log(1-s) \iff \mathbb{P}(L^*=n)=\frac{1}{n(n-1)},\ n=2,3,...$$

In chapter 4 we use the fixed-T finite dimensional formula of [24] to sketch prove that the L^*-tree has coalescent structure converging to the celebrated Bolthausen-Sznitman coalescent [6].

Theorem 1.2.10.

$$(\rho_t^{k,L^*,T})_{t\in[0,T]}\to^D(\bar{\rho}^{k,\mathsf{BS}})_{t\in[0,\infty)}$$

Where $(\rho_t^{k,\mathsf{BS}})_{t\in[0,\infty)}$ is the Bolthausen-Sznitman coalescent restricted to $\{1,...,k\}$.

Though this result is of interest on its own, our study of the collection $\{L^*_{1+\alpha},\alpha\in[0,1]\}$ is motivated by deep time-change results of Lageras and Sagitov [30]. In ongoing work, with Simon Harris, Juan Carlos Pardo Millán and Matt Roberts we use will use these time-changes to prove the theorems 1.2.11 and 1.2.12 about coalescent structures of heavy-tailed trees, though in chapter 4 we supply sketch proofs.

For $\alpha \in (0,1]$, we write $\mathcal{M}^{1+\alpha}$ for the critical offspring variables in the domain of attraction of a $(1+\alpha)$-stable law, which here we understand informally as

$$\mathcal{M}^{1+\alpha} = \left\{ L : \mathbb{E}[L] = 1, \forall \epsilon > 0, \mathbb{E}[L^{1+\alpha-\epsilon}] < \infty,\ \mathbb{E}[L^{1+\alpha+\epsilon}] = \infty \right\}. \tag{1.27}$$

Theorem 1.2.11. The trees with offspring distributions $L \in \mathcal{M}^{1+\alpha}$ form a universal class with the same limiting coalescent structure, in that there exists a limit process $(\pi_t^{k,\alpha-\mathsf{Crit}})_{t\in[0,1]}$ such that for every $L \in \mathcal{M}^{1+\alpha}$,

$$\left(\pi_{tT}^{k,L,T}\right)_{t\in[0,1]} \to^D \left(\bar{\pi}_t^{k,\alpha-\mathsf{Crit}}\right)_{t\in[0,1]} \equiv^{\mathsf{Law}} \left(\bar{\pi}^{k,L^*_{1+\alpha}}_{\log\left(\frac{1}{1-t}\right)}\right)_{t\in[0,1]} \tag{1.28}$$

where $\left(\bar{\pi}_t^{k,L^*_{1+\alpha}}\right)_{t\in[0,\infty)}$ is the limit process associated with the supercritical $L^*_{1+\alpha}$-tree. Furthermore, $\left(\bar{\pi}_t^{k,\alpha-\mathsf{Crit}}\right)_{t\in[0,1]}$ has splitting times given by

$$\mathbb{P}(\bar{\pi}^{k,\alpha-\mathsf{Crit}} : \eta_0 \prec^{dt_1} \eta_1 \prec^{dt_2} \dots \prec^{dt_n} \eta_n) \tag{1.29}$$

$$= \int_0^\infty \frac{\theta^{\alpha n}}{(k-1)!(1+\theta^\alpha)^{1+1/\alpha}} \prod_{i=1}^n \left\{ \frac{(1+\alpha)^{(c)}}{\alpha} \frac{1}{(1+\theta^\alpha(1-t_i))^c} dt_i \right\} dv. \tag{1.30}$$

Now we look at the case $\alpha = 0$, requiring a more delicate treatment of the tail asymptotics. For $\beta > 0$, again we give the informal definition

$$\mathcal{M}^1_\beta = \left\{ L : \mathbb{E}[L] = 1, \forall \epsilon > 0, \mathbb{E}[L(\log_+ L)^{\beta-\epsilon}] < \infty,\ \mathbb{E}[L(\log_+ L)^{\beta+\epsilon}] = \infty \right\}. \tag{1.31}$$

Theorem 1.2.12. For each $\beta > 0$, the trees with offspring distributions $L \in \mathcal{M}^1_\beta$ form a universal class whose coalescents converge to a β-dependent time-change of the Bolthausen-Sznitman coalescent. That is, for $L \in \mathcal{N}^1_\beta$,

$$\left(\rho_{tT}^{k,L,T}\right)_{t\in[0,1]} \to^D \left(\rho^{k,\mathsf{BS}}_{\frac{\beta}{\beta+1}\log\left(\frac{1}{1-t}\right)}\right)_{t\in[0,1]}.$$

1.3. Related work

A large portion of the existing literature concerned with coalescence in Galton-Watson trees has dealt in discrete time, so let us briefly discuss the similarities and differences between continuous-time and discrete-time Galton-Watson trees. Of particular interest to us is the notion of embeddability, a tool of conversion between the discrete and the continuous.

Suppose we have a continuous-time Galton-Watson process $(N_t)_{t\geq 0}$ with generating function $F_t(s)$. Consider that for $\delta > 0$, the process (N_t) viewed at lattice times

$$\{M_p = N_{p\delta} : p \in \mathbb{Z}_+\}$$

is a discrete-time Galton-Watson process $(M_p)_{p\in\mathbb{Z}_+}$ with offspring generating function $f(s) = F_\delta(s)$.

The *embeddability problem* is the task of classifying the set of discrete-time processes that can be constructed in this way - which amounts to identifying the 'embeddable' probability generating functions. We say a probability generating function $g(s) = \sum_{k=0} a_k s^k$ is embeddable if there exists a one-parameter semigroup of smooth transformations

$$\{F_t \in C^\infty([0,1],[0,1]) : t \geq 0\},\ F_u \circ F_v(s) = F_{u+v}(s)$$

such that $F_1(s) = g(s)$. See [5, Chapter III, Section 12] for further discussion.

The key idea for us is that, us is that when a coalescent formula holds in continuous-time, discrete-time analogues hold for all embeddable discrete trees simply by reading off continuous-time formula at integer times.

To give an idea of how this works, we offer an example. Pick two individuals from an embeddable discrete-time tree at a time T. Let $\tau^{\mathsf{disc}} \in \{0,1,2,...,T-1\}$ be the time at which they last shared a common ancestor. Since this discrete-time tree is embeddable, it can be embedded in a continuous-time tree. Let $\tau^{\mathsf{cts}} \in [0,T)$ be the time in the corresponding continuous-time tree at which they last shared a common ancestor. For integer $t \in \{0,1,...,T-1\}$ we have the relation

$$\left\{\tau^{L,T,\mathsf{cts}} \in [t,t+1)\right\} = \left\{\tau^{L,T,\mathsf{disc}} = t\right\}.$$

It should then come as no surprise to find that for instance, the discrete-time formula by Lambert (equation (1.1)) and Grosjean and Huillet (equation (1.32) below) agree with special cases of our continuous-time formula (1.14), replacing the continuous generating function $F_t(s)$ with the discrete analogue $f_n(s)$.

When a discrete tree is non-embeddable, there can be quite dramatic differences between a discrete tree and its most natural continuous-time analogues. Consider that the discrete-time binary tree is in fact a non-random object, and hence has less interesting structure than the Yule tree. If we pick uniformly two individuals from a discrete-time Yule-tree at time N, it is not difficult to show that the time τ at which they last shared their common ancestor is roughly Geometric with parameter $1/2$ as N gets large. Compare this with the more subtle distribution of the time to most recent common ancestor of the continuous-time Yule tree. This has density (1.4), a non-trivial probability density with tails decaying like te^{-t}.

Taking into consideration that continuous-time trees are often more natural models of real world processes, sometimes lend themselves more easily to analysis (for instance by the use of partial differential equations) and most importantly, that is easier to unify our work with the existing coalescent literature which operates mainly in continuous-time, we concentrate on continuous-time trees.

1.3.1 Coalescent Literature

Let $\mathcal{P}_k$ be the set of all partitions on $\{1,2,...,k\}$, and let $\mathcal{P}_\infty$ be the set of partitions on $\mathbb{N}$. For a partition $\rho \in \mathcal{P}_\infty$, and a subset $A \subseteq \mathbb{N}$, let ρ^A be the restriction of ρ to A.

A coalescent process is a process $(\rho_t)_{t\in[0,\infty)}$ taking values in $\mathcal{P}_k$ for some $k \in \mathbb{N} \cup \{\infty\}$, with the property that as time passes, blocks merge. That is,

$$s < t \implies \rho_s \succ \rho_t.$$

In his pioneering paper [42], Jim Pitman classified in their entirity the exchangeable and Markovian coalescents. The following definition of his famous Λ-*coalescents* is taken from the abstract.

> For each finite measure Λ on $[0, 1]$, a coalescent Markov process is constructed, so that the restriction of the partition to each finite subset of $\mathbb{N}$ is a Markov chain with the following transition rates: when the partition has b blocks, each k-tuple of blocks is merging to form a single block at rate
>
> $$\lambda_{b,k} = \int_0^1 x^{k-2}(1-x)^{b-k}\Lambda(dx).$$

Furthermore, Pitman proves that every process with rates given by $(\lambda_{b,k} : 0 \leq k \leq b)$ can be constructed in this way for some Λ.

The most notable coalescent process is Kingman's coalescent [28], characterised by unit-rate and pairwise mergers not depending on the number of total blocks. It is represented as a Λ-coalescent by setting $\Lambda = \delta_0$, and hence $\lambda_{b,k} = \mathbb{1}_{k=2}$. Kingman's coalescent commonly appears as the family tree for uniformly chosen individuals in constant population models, for example, in the classic Wright-Fisher model.

Another notable example is the Bolthausen-Sznitman coalescent [10], corresponding to the case where Λ is the Lebesgue measure, with merger rates

$$\lambda_{b,k} = \int_0^1 x^{k-2}(1-x)^{b-k}dx = \frac{(k-2)!(b-k)!}{(b-1)!}.$$

Bolthausen and Sznitman discovered this process in their analysis of Ruelle's probability cascades. It also makes appearences in the coalescent structure of Neveu's CSBP [9] and of branching Brownian motion on the half-line [7].

As we've seen so far, in general, the processes $(\pi_t^{k,L,T})_{t\in[0,T]}$, $(\rho_t^{k,L,T})_{t\in[0,T]}$, and even their T-limits, are non-Markovian in general. However, when the population is large at the picking time, often there exists some notion of exchangability or fairness in both the splits of the process (π_t^T) and the mergers of the coalescent (ρ_t^T). With this, it is natural that in some sense that topological propertys from Pitman's Λ-coalescents are inherited. In particular, we've seen the processes $\left(\bar{\pi}_t^{k,\mathsf{Crit}^\mu}\right)_{t\in[0,1]}$ share topological structure with Kingman's coalescent.

1.3.2 Coalescence in Galton-Watson trees

Several authors have studied coalescence in Galton-Watson trees. Where possible, we convert the results of other authors into our notation. We write N and n in place of T and t, and occasionally a superscript disc, to emphasise a result is concerned with discrete-time. For discrete time processes $(Z_n)_{n\in\mathbb{Z}_+}$ we write $f(s)$ for the offspring generating function and $f_n(s) = \mathbb{E}[s^{Z_n}]$ for the n^{th} iterate.

In [17], Grosjean and Huillet examine the coalescent structure of discrete-time Galton-Watson processes starting with $x \in \{1, 2, ...\}$ individuals, providing various formula associated with the (also discrete-time) coalescent process $(\rho_n^{k,L,N})_{n\in\{0,1,...,N\}}$ with a focus on the time to the most recent common ancestor of all k individuals. In particular, let $\upsilon^{k,L,N}$ be the different between $N - \tau^{k,L,N}$, where $\tau^{k,L,N}$ is the time of the most recent common ancestor of all k individuals. If not all k individuals are descendended from the same time-0 ancestor, set $\upsilon^{k,L,N} = \infty$. Grosjean and Huillet provide the following generalisation of Lambert's formula (1.1).

Lemma 1.3.1 (Grosjean and Huillet [17], Corollary 7)**.** For $N \geq 1$ and $1 \leq n \leq N$,

$$\mathbb{P}(\upsilon^{k,L,N} \leq n) = x \int_0^1 (1-s)\frac{f_n^k(s)}{f_n'(s)} f_N'(s) f_N(s)^{x-1} ds \tag{1.32}$$

In his wide spanning paper [31], Amaury Lambert considered the coalescent structure of discrete-time Galton-Watson processes and continuous-state branching processes. As we saw in the introduction, he gave (1.32) in the case $k = 2$, and went on to study the asymptotics of the time to most recent common ancestor in the subcritical Galton-Watson trees. Pick two individuals from a subcritical tree at time N and let $\upsilon^{x,L,N}$ be the time to most recent common ancestor from time N. Lambert established the distributional convergence

$$\upsilon^{x,L,N} \to \upsilon^{L,\mathsf{disc}} \in \{1, 2, ...\}, \tag{1.33}$$

where the limit variable $\upsilon^{L,\mathsf{disc}}$ doesn't depend on x, and has implicit characterisation

$$\mathbb{E}^{\mathsf{qs}}\Big[\tilde{Z}(\tilde{Z}-1)s^{\tilde{Z}-2}; \upsilon^{L,\mathsf{disc}} \leq n | \tilde{Z} \geq 2] = \frac{1}{1-g'(0)} g'(s) \frac{f_n''(s)}{f_n'(s)}, \tag{1.34}$$

where $\tilde{Z}$ is the limiting value of Z_t conditional on $Z_t \geq 2$ as $t \to \infty$, and g is its generating function (so that hence $1 - g'(0) = \mathbb{P}(\tilde{Z} \geq 2)$).

In [36], Le replicates in continuous-time the Lambert's results of [31]. Le also considered the coalescence structure of larger samples of individuals, giving an implicit representation for the laws of binary coalescent trees. More specifically, Le provided an implicit representation for the law of the event

$$A := \{\eta_0 \to^{dt_1} \eta_1 \to^{dt_2} ... \to^{dt_{n-1}} \eta_n\}$$

in the special case that each η_i is obtained from η_{i-1} by a block breaking into precisely two blocks, that is, the ancestral tree associate with the process $(\pi_t^{k,T})_{t\in[0,T]}$ is a binary tree. Le showed

$$\mathbb{E}_x[N_T^{(k)} s^{N_T - k}; A, N_T \geq k] = xF_T'(s)F_T(s)^{x-1} \prod_{i=1}^{k-1} F_{T-t_i}'(s) f''(F_{T-t_i}(s)), \tag{1.35}$$

where $N_T^{(k)} = N_T(N_T-1)...(N_T-k+1)$. The omission of analogous formula for the general (non-binary) tree appears to be an oversight.

We now provide an inversion formula that allows us to express (1.34) and (1.35) as an integral formula depending only on the generating functions of the process, showing that they agree with special cases (1.16).

Lemma 1.3.2 (Inversion formula)**.** Suppose N is a random variable and A is an event on some probability space, and that

$$\mathbb{E}[N^{(k)}s^{N-k};A] = \Lambda_A(s).$$

Then the conditional law of A can be recovered by the inversion

$$\mathbb{P}(A|N \geq k) = \int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N \geq k)}\Lambda_A(s)ds$$

Proof. It is a straightforward consequence of beta integrals that for an integer $n \in \{k, k+1, ...\}$,

$$\frac{1}{n^{(k)}} = \frac{(n-k)!}{n!} = \frac{1}{(k-1)!}\int_0^1 (1-s)^{k-1}s^{n-k}ds.$$

Using Fubini in the third equality below, it follows that

$$\begin{aligned}
\int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N_T \geq k)}\Lambda_A(s)ds &= \int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N \geq k)}\mathbb{E}[N^{(k)}s^{N-k};A]ds \\
&= \int_0^1 \frac{(1-s)^{k-1}}{(k-1)!}\mathbb{E}[N^{(k)}s^{N-k};A|N \geq k]ds \\
&= \mathbb{E}\left[\frac{N^{(k)}}{(k-1)!}\int_0^1 (1-s)^{k-1}s^{N-k}ds; A\middle|N \geq k\right] \\
&= \mathbb{P}(A|N \geq k).
\end{aligned}$$

□

Applying this to Lambert's formula (1.34) for subcritical time to most recent common ancestors, we yield

$$\mathbb{P}(v^{L,\mathsf{disc}} \leq n) = \int_0^1 \frac{(1-s)}{1-g'(0)}g'(s)\frac{f_n''(s)}{f_n'(s)}ds. \tag{1.36}$$

Replacing n with t, and differentiating, this agrees with our subcritical formula (1.23) in the case $k=2$.

Applying this lemma to Le's formula (1.35), we see that

$$\mathbb{P}_x(\eta_0 \to^{dt_1} \eta_1 \to^{dt_2} ... \to^{dt_{n-1}} \eta_n) = \int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N_T \geq k)}xF_T'(s)F_T(s)^{x-1}\prod_{i=1}^{k-1}F_{T-t_i}'(s)f''(F_{T-t_i}(s))ds$$

Setting $x=1$, this agrees with the split time representation (1.16) of the law of the partition process $(\pi_t^{k,L,T})_{t\in[0,T]}$ in the special case $(c_1, ..., c_n) = (\underbrace{2, 2, ..., 2}_{k-1 \text{ times}})$, giving the density of split times for binary trees.

1.3.3 Coalescence in CSBPs

Several authors have considered coalescent questions related to continuous-state branching processes (CSBP). In short, a Markov process $(Y_t : t \geq 0)$ with probabilities $\{\mathbb{P}_x : x \geq 0\}$ is called a continuous state branching processes if its law observes the branching property:

$$\mathbb{E}_x[e^{-\theta Y_t}] = e^{-xu_t(\theta)},\ \forall \theta \geq 0, \forall t \geq 0, \tag{1.37}$$

for some function $u_t(\theta)$ characterised by a branching mechanism ψ through the equation

$$\frac{\partial u_t}{\partial t}(\theta) + \psi(u_t(\theta)) = 0.$$

We refer the reader to Kyprianou [29, Chapter 12] for details.

In [9], Bertoin and Le-Gall provide a ingenious construction making rigorous intuitive notions of genealogy in continuous-state branching processes. They construct a doubly-indexed process

$$\{Y(t,a) : t \geq 0, a \geq 0\}$$

where $Y(\cdot, 0) = 0$, and for each a, b, $Y(\cdot, a+b) - Y(\cdot, a)$ is independent of the family of processes $\{Y(\cdot, c) : c \leq a\}$ and is distributed like a CSBP with branching mechanism ψ and starting at b. The following proposition makes resourceful use of the Levy structure inherent in looking at a CSBP 'on its side'.

Proposition 1.3.3 ([9], Proposition 1)**.** There exists a process $\{S^{(s,t)}(a) : 0 \leq s \leq t, a \geq 0\}$ such that

1. For every $s \leq t$, $(S^{(s,t)}(a) : a \geq 0)$ is a subordinator with Laplace exponent $u_{t-s}(\cdot)$.
2. For every integer $p \geq 2$, and $0 \leq t_1 \leq ... \leq t_p$, the subordinators $S^{(t_1,t_2)}, ..., S^{(t_{p-1},t_p)}$ are independent, and

$$S^{(t_1,t_p)}(a) = S^{(t_{p-1},t_p)} \circ \ldots \circ S^{(t_1,t_2)}(a). \tag{1.38}$$

Finally, the process $\{S^{(0,t)}(a) : t \geq 0, a \geq 0\}$ and $\{Y(t,a) : t \geq 0, a \geq 0\}$ have the same finite dimensional marginals.

This proposition can be used to devise a precise notion of ancestry in a CSBP. For $t < T$ we say an individual b at time t is an ancestor of individual c at time T if

$$S^{(t,T)}(b-) < c < S^{(t,T)}(b). \tag{1.39}$$

We define the random partition process $(\pi_t^{k,a,\psi,T})_{t\in[0,T]}$ associated with a CSBP as follows. Let $\{Y_t : t \geq 0\}$ be a CSBP starting at a with branching mechanism ψ. On the event $\{Y_T > 0\}$, sample k points $C_1, ..., C_k$ independently and uniformly on the interval $[0, Y_T]$. We let $\pi_t^{k,a,\psi,T}$ be the partition of $\{1, ..., k\}$ associated with the equivalence relation

$$i \sim_t k \iff \exists b \in [0, Y_t] : C_i, C_j \in (S^{(t,T)}(b-), S^{(t,T)}(b))$$

Setting $\rho_t := \pi_{T-t}$, Bertoin and Le Gall study the law of the process $(\rho_t^{k,a,\psi,T})_{t\in[0,T]}$ when the branching mechanism is given by

$$\psi^{\mathsf{Nev}}(\lambda) = \lambda \log(\lambda). \tag{1.40}$$

The ψ^{Nev}-CSBP is known as Neveu's CSBP (see [37]). Remarkly, the law of $(\rho_t^{k,a,\psi,T})_{t\in[0,T]}$ doesn't depend on a, and furthermore,

$$(\rho_t^{k,a,\psi,T})_{t\in[0,T]} \equiv^{\mathsf{Law}} (\rho_t^{\mathsf{k,BS}})_{t\in[0,T]},$$

where $\rho^{k,\mathsf{BS}}$ is the Bolthausen-Sznitmann coalescent restricted to $\{1,...,k\}$.

Lambert used this construction in[31] to tackle the following problem for general ψ. Suppose we have a CSBP $(Z_t)_{t\geq 0}$ starting at $x>0$ and with branching mechanism ψ, and condition on the event $Z_T>0$. What is the distribution of $\tau^{x,\psi,T}$, the time at which two uniformly chosen individuals at time T last shared a common ancestor?

Lambert ultimately proves the following result.

Lemma 1.3.4 ([31], Corollary 2)**.**

$$\mathbb{P}(\tau^{x,\psi,T}\in[t,T]) = \int_0^\infty \lambda \frac{u''_{T-t}(\lambda)}{u'_{T-t}(\lambda)} u'_t(\lambda) e^{-xu_t(\lambda)} d\lambda$$

Lambert then considers asymptotics $\tau^{x,\psi,T}$ as $T\to\infty$ in the case where the CSBP is critical or subcritical. Notably, in the case where the CSBP is a critical Feller diffusion, that is, governed by branching mechanism

$$\psi^{\mathsf{Fel}}(\lambda) = \frac{1}{2}\lambda^2$$

he establishes the distributional convergence

$$\tau^{x,\psi^{\mathsf{Fel}},T}/T \to^{D} \bar{\tau}^{\mathsf{Crit}} \in [0,1]$$

where $\bar{\tau}^{\mathsf{Crit}}$ doesn't depend on x, and is the same random variable we saw in the Galton-Watson case, with law given by (1.8). Lambert also considers the same problem for a Feller diffusion with drift: $\psi(\lambda)=\mu\lambda+\frac{1}{2}\lambda^2$, and obtains the random variable $\tau^{\mathsf{crit}^{\mu}}$ seen in (1.9).

The underlying reasons for the universality of the random variable $\bar{\tau}^{\mathsf{Crit}}$ are natural. Feller first obtained his epynomous diffusion by studying rescalings of critical Galton-Watson processes, see [41] for a recent article surveying this connection. Evidently the large-T asymptotics for coalescent times within these processes commute with this rescaling.

1.3.4 Coalescence in random trees

In [35], Lambert and Stadler consider the coalescent structure of a large class of binary trees called coalescent point processes. If one takes an independent Bernoulli sample with probability $p\in(0,1]$ of the population at time T, they provide formula for the coalescent structure of their ancestral subtree. In [32], Lambert shows that the contour process associate with splitting trees is a Levy process, and utilises this Lévy structure to provide most recent common ancestor formula for neighbouring leaves in the splitting tree.

In [43], Lea Popovic makes great use of Brownian excursion theory to study the coalescent structure of the entire population in critical binary branching processes, deriving a point process interpretation of the simultaneous genealogy of the entire population. In [2], Popovic and Aldous generate a formula identical to (1.25) as an answer to a completely question. Namely, they consider critical birth-death trees started from an improper random time in the past, finding that after rescaling, these trees are identical in law to $(\bar{\pi}_t^{k,\mathsf{Crit}})_{t\in[0,1]}$.

1.4. Structure of thesis

- Chapter 2 comprises the paper [24], establishing fixed-T formula for the law of the process $(\pi_t^{k,L,T})_{t\in[0,T]}$, as well as distributional convergence in the supercritical and subcritical cases.
- Chapter 3 comprises the paper [20], which gives two main results. The first is an explicit computation of the law of $(\pi_t^{k,L,T})_{t\in[0,T]}$ when $\mathbb{P}(L \in \{0,2\}) = 1$. The second result shows that a large class of near-critical processes have coalescent structures converging to time changes of a limit process topologically equivalent to Kingman's coalescent.
- In chapter 4 we study the coalescent structure of heavy-tailed trees. In particular, we prove that the L_*-coalescent converges to Bolthausen-Sznitman, and we sketch prove theorems about the limiting coalescent structures of heavy-tailed trees.

Chapter 2

Coalescence at fixed times and supercritical and subcritical trees.

This chapter consists of the paper ‘Coalescence in continuous-time Galton-Watson trees’ [24], submitted to arXiv (arXiv:1709.08500) in September 2017.

The coalescent structure of supercritical and subcritical continuous-time Galton-Watson trees

S. G. G. Johnston

School of Mathematics and Statistics, University College Dublin, Belfield, Dublin 4

January 26, 2018

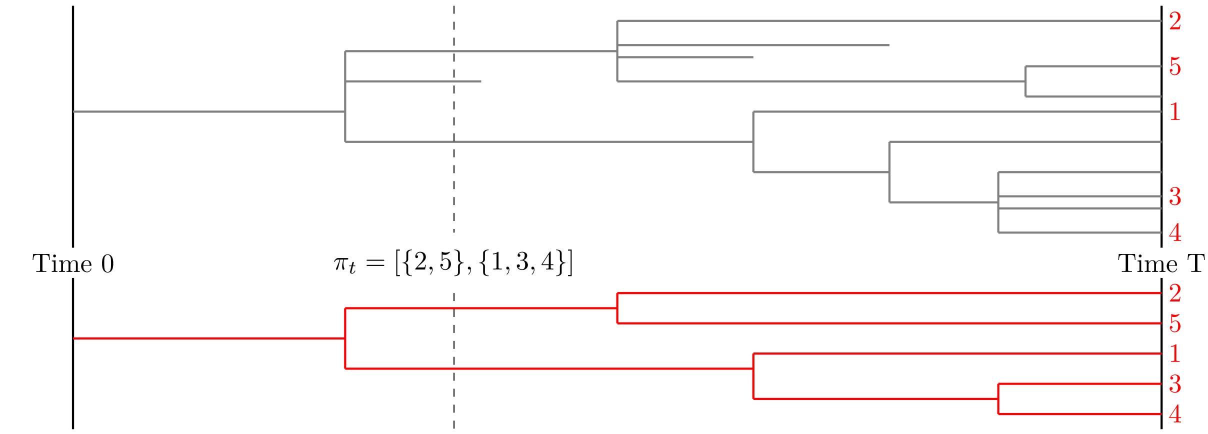
Abstract

Take a continuous-time Galton-Watson tree with offspring distribution L. On the event that there are at least k particles alive at a fixed time T, choose k distinct particles uniformly from those alive. For each time $t \in [0,T]$, we can create a random partition $\pi_t^{k,L,T}$ of $\{1, ..., k\}$ by $i \sim_{\pi_t} j$ if i and j share a common ancestor alive at time t. Using spine methods, we give formulas characterising the law of the process $(\pi_t^{k,L,T})_{t \in [0,T]}$, and study the asymptotics of these formulas in supercritical and subcritical trees to establish the convergence of $(\pi_t^{k,L,T})_{t \in [0,T]}$ to particular limit processes as $T \to \infty$. These asymptotic results complement the work of Harris, Johnston and Roberts [20], who looked at the critical case.

2.1. Introduction

Let L be a random variable taking values in $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$, and let $f(s) = \mathbb{E}[s^L]$ be its generating function. Consider a continuous-time Galton-Watson tree branching at rate 1 with offspring distribution L in which we begin with one initial particle. Let N_t be the number of particles alive at time t, and let $F_t(s) = \mathbb{E}[s^{N_t}]$ be the generating function of the process.

For $T > 0$ we condition on the event $\{N_T \geq k\}$, and pick k particles alive at time T uniformly and without replacement. Label these particles $1, 2, ..., k$, and consider their ancestral tree.



For each time $t \in [0, T]$, we define the equivalence relation $i \sim_{\pi_t} j$ if and only if i and j share a common ancestor alive at time t. We let $\pi_t^{k,L,T}$ denote the random partition of $\{1, ..., k\}$ corresponding to this equivalence relation. The process $(\pi_t^T) := (\pi_t^{k,L,T})_{t \in [0,T]}$ is a right-continuous partition-valued process characterising the entire ancestral tree of the k particles labelled at time T. In many cases, it is more natural to consider the process $(\rho_t^T) := (\rho_t^{k,L,T})_{t \in [0,T]}$ defined by $\rho_t^{k,L,T} = \pi_{T-t}^{k,L,T}$. Our goal is to describe the law of (π_t^T) and (ρ_t^T), with a focus on the asymptotic regime $T \to \infty$.

Before stating our main results, we discuss the case $k = 2$, for which some results are already known. This case amounts to choosing two individuals from a tree with offspring distribution L at a time T, and studying the time $\tau^{L,T} \in [0, T]$ at which they last shared a common ancestor.

In this direction, the following result (which we will generalise significantly later) was proved by Lambert [31] in discrete-time. Le gave the continuous-time analogue in [36].

Lemma 2.1.1 (Lambert, 2003. Le, 2014)**.** Conditioned on $\{N_T \geq 2\}$, pick uniformly and without replacement two particles of those alive at time T, and label the time at which they shared a common ancestor $\tau^{L,T}$. Then

$$\mathbb{P}\Big(\tau^{L,T} \in [t, T]\Big) = \frac{1}{\mathbb{P}(N_T \geq 2)} \int_0^1 (1-s) \frac{F''_{T-t}(s)}{F'_{T-t}(s)} F'_T(s) ds$$

where we recall $F_t(s) = \mathbb{E}[s^{N_t}]$ is the generating function for the number alive in the process.

Although this result gives a powerful implicit characterisation of the distribution of $\tau^{L,T}$, it is difficult to infer qualitative properties of this random variable directly from the formula above. By sending $T \to \infty$, however, it is possible to gain a more intuitive insight. Unsurprisingly, different qualitative behaviours arise depending on whether the underlying Galton-Watson process is supercritical, critical, or subcritical. Letting $m := \mathbb{E}[L]$ denote the mean number of offspring, these cases correspond to $m > 1$, $m = 1$, and $m < 1$ respectively.

When the process is supercritical, the time $\tau^{L,T}$ remains near the beginning of the interval $[0, T]$, even when T is very large:

Lemma 2.1.2. Suppose $\mathbb{E}[L] > 1$ and the Kesten-Stigum condition $\mathbb{E}[L \log_+ L] < \infty$ holds. Then we have the following convergence in distribution as $T \to \infty$:

$$\tau^{L,T} \xrightarrow{D} \bar{\tau}^L \in [0, \infty).$$

Furthermore, the probability density of the limit variable $\bar{\tau}^L$ is given by the integral representation

$$\mathbb{P}(\bar{\tau}^L \in dt) = e^{-(m-1)t} \int_0^\infty \frac{v\varphi^1(v)}{1 - \varphi^0(\infty)} \varphi^1(ve^{-(m-1)t}) f''\Big(\varphi^0(ve^{-(m-1)t})\Big) dv, \; t \in [0, \infty) \quad (2.1)$$

where $\varphi^j(v) = \mathbb{E}[W^j e^{-vW}]$, $W = \lim_{t\to\infty} N_t e^{-(m-1)t}$ is the martingale limit, and $(1 - \varphi^0(\infty))$ is the survival probability.

The convergence in distribution of $\tau^{L,T}$ was proved by Athreya in [3]. The distributional characterisation (2.1) is new, and is a special case of our main supercritical Theorem, 2.2.4.

By way of example, an explicit formula is available when $L = 2$ almost surely:

Example 2.1.1. We call our process a standard Yule tree if $f(s) = s^2$. In this case, it is well known that the martingale limit W is a unit mean exponential, and hence

$$f''(u) \equiv 2, \; \varphi^1(v) = \frac{1}{(1+v)^2}. \quad (2.2)$$

Inserting this into (2.1), a calculation shows that the limit variable $\bar{\tau}^{\mathsf{Yule}}$ has an interesting and nontrivial probability density function

$$\mathbb{P}(\bar{\tau}^{\mathsf{Yule}} \in dt) = \frac{2e^t}{(e^t - 1)^3} \Big[(t-2)e^t + (t+2)\Big] dt, \; t \in [0, \infty). \quad (2.3)$$

Moving on to the critical case, the following result shows that whenever $\mathbb{E}[L^2] < \infty$, the variable τ^T is stretched proportionally to T:

Lemma 2.1.3. Suppose $\mathbb{E}[L] = 1$ and $\mathbb{E}[L^2] < \infty$. Then we have the following convergence in distribution as $T \to \infty$:

$$\frac{\tau^{L,T}}{T} \xrightarrow{D} \bar{\tau}^{\mathsf{Crit}} \in [0, 1], \quad (2.4)$$

and the limit variable $\bar{\tau}^{\mathsf{Crit}}$ is universal in all critical L with $\mathbb{E}[L^2] < \infty$. Furthermore, the limit variable $\bar{\tau}^{\mathsf{Crit}}$ has probability density function

$$\mathbb{P}(\bar{\tau}^{\mathsf{Crit}} \in dt) = \Big[-4\frac{\log(1-t)}{t^3} + 2\frac{\log(1-t)}{t^2} - 4\frac{1}{t^2}\Big] dt, \; t \in [0, 1]. \quad (2.5)$$

This result is known. See Durrett [13], O'Connell [39], Lambert [31], Athreya [3], and Harris, Johnston, and Roberts [20].

In the subcritical case, survival is extremely hard. As a result, if the process survives until a large time T, it is overwhelmingly likely that just one ancestral lineage survived for most of the interval $[0, T]$. Indeed, the following result says that the difference between T and $\tau^{L,T}$ converges:

Lemma 2.1.4. In the subcritical case, the following convergence holds as $T \to \infty$:

$$T - \tau^{L,T} \xrightarrow{D} \bar{v}^L \in [0, \infty). \tag{2.6}$$

Furthermore, the subcritical limit variable $\bar{v}^L$ has probability density given by

$$\mathbb{P}(\bar{v}^L \in dt) = \int_0^1 \frac{(1-s)}{\mathbb{P}(W \geq 2)} B'(s) F_t'(s) f''(F_t(s)) ds, \ t \in [0, \infty). \tag{2.7}$$

where W is the quasi-stationary limit variable:

$$\mathbb{P}(W = n) = \lim_{t \to \infty} \mathbb{P}(N_t = n | N_t \geq 1),$$

and $B(s) = \mathbb{E}[s^W]$ is the generating function of the quasi-stationary limit.

This result can be found in [36] and [31] and, but is the special case $k = 2$ of Theorem 2.2.6. See also Athreya [3].

Let us now give a brief overview of our main results, which will be stated formally in the next section. First we will study the process $(\pi_t^T) := (\pi_t^{k,L,T})_{t \in [0,T]}$ for fixed (finite) T. Theorems 2.2.2 and 2.2.3 characterise the law of this process in two different ways, first in terms of its finite dimensional distributions and then in terms of the random splitting times of blocks in (π_t^T). In both cases, explicit formulas are obtained, each in the form of an integral equation involving various generating functions associated with the process.

We then send the picking time $T \to \infty$, and study the asymptotic behaviour of the process (π_t^T). As in the case $k = 2$ discussed in the introduction, we will see analogous qualitative differences depending on the mean of the offspring distribution. Namely, in the supercritical case the common ancestors of a sample of k individuals picked at a large time T last existed near the start of $[0, T]$, and in the subcritical case, the common ancestors last existed near the end of $[0, T]$. The critical case has already been covered in [20], so in the interest of completeness we mention very briefly the results from this paper below.

First we will consider the supercritical case, where $m := \mathbb{E}[L] > 1$. Under the Kesten-Stigum condition, $\mathbb{E}[L \log_+ L] < \infty$, we will see in Theorem 2.2.4 that the process $(\pi_t^{k,L,T})_{t \in [0,T]}$ satisfies the distributional convergence

$$(\pi_t^{k,L,T})_{t \in [0,T]} \xrightarrow{D} (\bar{\pi}_t^{k,L})_{t \in [0,\infty)}, \tag{2.8}$$

and we will characterise the law of the limit process $(\bar{\pi}_t^{k,L})_{t \in [0,\infty)}$ in terms of the generating function of the martingale limit

$$\varphi(v) = \mathbb{E}[e^{-vW}] := \lim_{t \to \infty} \mathbb{E}[e^{-vN_t e^{-(m-1)t}}].$$

In the subcritical case, we will see that the subcritical partition process satisfies

$$(\rho_t^{k,T,L})_{t\in[0,T]} \xrightarrow{D} (\bar{\rho}_t^{k,L})_{t\in[0,\infty)}, \tag{2.9}$$

and we will characterise the law of $(\bar{\rho}_t^{k,L})_{t\in[0,\infty)}$ in terms of the quasi-stationary generating function

$$B(s) = \mathbb{E}[s^W] := \lim_{t\to\infty} \mathbb{E}[s^{N_t}|N_t \geq 1].$$

The rest of the paper is structured as follows. In section 2, we outline our main results. In section 3, we introduce the idea of multiple spines, privileged lines of descent that flow through the tree. We then create a change of measure which forces the spines to represent a uniform sample of k individuals at time T. In section 4, we prove the formula of section 2.1 characterising the distribution of (π_t^T). In section 5, we then give proofs of our asymptotic formulas, proving the results of section 2.2.

2.2. Main Results

Before stating our main results, we need to introduce some more notation and two key hypotheses. We start by giving a brief formal description of the continuous time Galton-Watson tree. Let L be a $\{0,1,2,...\}$-valued random variable. Under the probability measure $\mathbb{P}$, we start at time 0 with one particle which we call $\varnothing$. The particle $\varnothing$ lives for an exponentially distributed, rate 1, length of time $\tau_\varnothing$ until it dies, and is replaced by a random number of offspring with labels $1,2,...,L_\varnothing$, where $L_\varnothing$ is distributed like L and is independent of $\tau_\varnothing$. These offspring then independently repeat this behaviour. That is, for each u born at some time, u lives a length of time τ_u distributed like $\tau_\varnothing$ and at death is replaced by offspring with labels $u1, u2, ..., uL_u$, where L_u is distributed like L. Here, τ_u and L_u are independent of each other and of the past. We write $\mathcal{N}_t$ for the set of particles alive at time t and $N_t = |\mathcal{N}_t|$ for the number of particles alive. We also write $u < v$ if u is an ancestor of v, and $u \leq v$ if $u < v$ or $u = v$.

A partition γ of $\{1,...,k\}$ is a collection of disjoint non-empty subsets $\Gamma_1,...,\Gamma_m$ of $\{1,...,k\}$, or'blocks', whose union is $\{1,...,k\}$. Equivalently, a partition is an equivalence relation on $\{1,...,k\}$. We will always order the blocks Γ_i of a partition γ by order of least element. We write $\mathcal{P}_k$ for the collection of partitions of $\{1,...,k\}$. For $\gamma = [\Gamma_1,...,\Gamma_p] \in \mathcal{P}_k$, we write $|\gamma| = p$ for the number of blocks in the partition. For partitions α,β, we say α can break into β, written $\alpha \prec \beta$, if β can be obtained by breaking up the blocks of α. For example, $\big[\{1,2,4\},\{3\}\big] \prec \big[\{1\},\{2,4\},\{3\}\big]$. Since $\alpha \prec \alpha$, the pair $(\mathcal{P}_k, \rightarrow)$ forms a partially ordered set. A chain of partitions is a sequence $[\{1,...,k\}] = \gamma_0, ..., \gamma_{n+1} = [\{1\},...,\{k\}]$ starting at the one block partition and finishing with the partition of $\{1,...,k\}$ into singletons, with the property that $\gamma_i \prec \gamma_{i+1}$ for every i. We emphasise that consecutive terms in a chain may be equal, a departure from the usual terminology in the the theory of partially ordered sets. Let $\mathcal{CP}_k^n$ be the set of chains $(\gamma_0,\gamma_1,...,\gamma_n,\gamma_{n+1})$ with entries in $\mathcal{P}_k$ of internal length n.

For a chain $(\gamma_0,\gamma_1,...,\gamma_n,\gamma_{n+1}) \in \mathcal{CP}_k^n$, we will write $\Gamma_{i,j}$ for the jth block (ordered by least element) of γ_i. By definition, each $\Gamma_{i,j}$ is the union of some classes in γ_{i+1}. With this in mind, we write $\gamma_{i,j}$ for the restriction of γ_{i+1} to $\Gamma_{i,j}$. We call $b_{i,j} = |\gamma_{i,j}|$ the breakage numbers of the partition sequence. That is, $b_{i,j} \geq 1$ is the number of blocks the jth block of γ_i needs to break into to form some new blocks in γ_{i+1}.

Note that since our partition chains always start with γ_0, comprising the single block $\{1,...,k\}$, the breakage number $b_{0,1}$ of the single block in γ_0 is equal to the number of classes in γ_1. Similarly, since our partition chains always end with γ_{n+1}, the partition of singletons, the breakage numbers of the blocks in γ_n, $b_{n,1},...,b_{n,|\gamma_n|}$ are precisely the sizes of the blocks in γ_n, that is $(b_{n,1},...,b_{n,|\gamma_n|}) = \big(|\Gamma_1|,....,|\Gamma_{n,|\gamma_n|}|\big)$.

For k distinct particles alive in a branching process at time T with labels $1,...,k$, let π_t be the partition generated by the equivalence relation $i \sim_{\pi_t} j$ if and only if i and j share a common ancestor alive at time t. Note that $(\pi_t)_{t\in[0,T]}$ is then a right-continuous process taking values in $\mathcal{P}_k$, and $\pi_s \prec \pi_t$ for $s \leq t$, and in particular,

$$\pi_0 = \Big[\{1,2,....,k\}\Big], \pi_T = \Big[\{1\},\{2\},...,\{k\}\Big].$$

Let $(\rho_t^{k,L,T})_{t\in[0,T]}$ be the process defined by by $\rho_t^{k,L,T} = \pi_{T-t}^{k,L,T}$. (ρ_t^T) is a coalescent process, in the sense that blocks merge together as time progresses.

Finally, we need to ensure that there actually *are* at least k particles alive at time T with positive probability, and that we can choose uniformly from them. To be more precise, we must ensure that both $\mathbb{P}(N_T \geq k) > 0$, and $P(N_T < \infty) = 1$. The inequality $\mathbb{P}(N_T \geq k) > 0$ is guaranteed to hold by virtue of our first hypothesis, which states that

$$P(L \geq 2) > 0. \tag{2.10}$$

In addition to (2.10), we insist that the following non-explosion hypothesis holds:

$$\int_{1-\epsilon}^{1} \frac{ds}{|f(s)-s|} = \infty \qquad \forall \epsilon \text{ near } 0. \tag{2.11}$$

This condition is equivalent to our second requirement that $\mathbb{P}(N_t < \infty) = 1$ for t, and is weaker than $\mathbb{E}[L] < \infty$. See [18, Chapter II, Theorem 9.1] for details. We emphasize that both hypotheses (2.10) and (2.11) are in force in the remainder of this paper.

We are now ready to state our main results, which we split into two sections. The results in section 2.2.1 concern fixed and finite T. The results in section 2.2.2 concern the asymptotic regime in which T is sent to ∞.

2.2.1 Law of the partition process $(\pi_t)_{t\in[0,T]}$

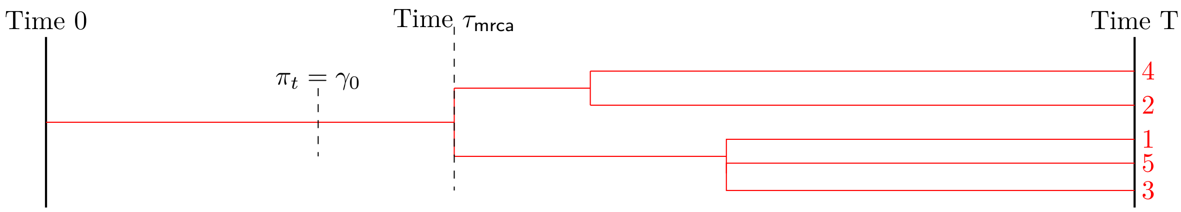
On the event $\{N_T \geq k\}$, pick uniformly k distinct particles alive at time T and label them $1,....,k$. Let $(\pi_t^T) = (\pi_t^{k,L,T})_{t\in[0,T]}$ be their partition process. Our main theorems in this section describe the law of the process (π_t^T) in terms of the generating functions $F_t(s) = \mathbb{E}[s^{N_t}]$ and $f(s) = \mathbb{E}[s^L]$. Below we will write $F_t^j(s)$ for the j^{th}-derivative of $F_t(s)$ with respect to s.

We start by calculating the one-dimensional distributions of the process $(\pi_t^{k,L,T})_{t\in[0,T]}$. That is, for fixed $t \in [0,T]$ we give the law of the $\mathcal{P}_k$-valued random variable $\pi_t^{k,L,T}$.

Theorem 2.2.1. For a partition γ of $\{1,...,k\}$ into p blocks $\Gamma_1,...,\Gamma_p$ of sizes $k_1+...+k_p = k$,

$$\mathbb{P}(\pi_t^{k,L,T} = \gamma) = \int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N_T \geq k)}\ F_t^p(F_{T-t}(s)) \prod_{i=1}^{p} F_{T-t}^{k_i}(s)ds. \tag{2.12}$$

Example 2.2.1 (The most recent common ancestor of k individuals)**.** By taking $\gamma = \gamma_0 := [\{1, ..., k\}]$ to be the partition of $\{1, ..., k\}$ into one block, we get a formula for the time to most recent common ancestor $\tau_{\mathsf{mrca}}^{k,L,T} \in [0, T]$. It should be clear from the diagram below that $\{\tau_{\mathsf{mrca}} > t\} \equiv \{\pi_t = \gamma_0\}$.



Hence by plugging in $\gamma = \gamma_0$ into equation (2.12),

$$\mathbb{P}(\tau_{\mathsf{mrca}}^{k,L,T} > t) = \mathbb{P}(\pi_t^{k,L,T} = \gamma_0) = \frac{1}{(k-1)!\mathbb{P}(N_T \geq k)} \int_0^1 (1-s)^{k-1} F_t^1(F_{T-t}(s)) F_{T-t}^k(s). \tag{2.13}$$

To see that equation (2.13) reduces to Lemma 2.1.1 in the case $k = 2$, note that by the branching property, the generating function satisfies the semigroup identity

$$F_u \circ F_v(s) = F_{u+v}(s),$$

and hence $F_T'(s) = F_{T-t}'(s) F_t'(F_{T-t}(s))$. We discuss the semigroup identity in further detail below.

More generally, the following theorem gives the finite dimensional distributions of (π_t^T).

Theorem 2.2.2 (Finite dimensional distributions of (π_t^T))**.** Let $0 = t_0 < t_1 < ... < t_n < t_{n+1} = T$, and write $\Delta t_i = t_{i+1} - t_i$. Let $[\{1, ..., k\}] = \gamma_0 \prec \gamma_1 \prec \gamma_2 ... \prec \gamma_n \prec \gamma_{n+1} = [\{1\}, \ldots, \{k\}]$ be a chain of partitions with breakage numbers b_{ij}. Then

$$\mathbb{P}(\pi_{t_1} = \gamma_1, ..., \pi_{t_n} = \gamma_n) = \int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N_T \geq k)} \prod_{i=0}^{i=n} \prod_{\Gamma_{i,j} \in \gamma_i} F_{\Delta t_i}^{b_{i,j}} \Big(F_{T-t_{i+1}}(s) \Big) ds. \tag{2.14}$$

An alternate way to characterise the law of (π_t^T) is in terms of the 'split times', the times t at which π_{t-}^T differs from π_t^T. Let $[\{1, ..., k\}] = \eta_0 \prec \eta_1 \prec ... \prec \eta_n = [\{1\}, ..., \{k\}]$ be a chain of partitions that is maximal in the sense that η_i is created from η_{i-1} by breaking precisely one block of η_{i-1} into $c_i \geq 2$ blocks in η_i. Note that $1 \leq n \leq k-1$, and that the set of multiplicities $\{c_i : i = 1, ..., n\}$ satisfy the branch equation

$$(c_1 - 1) + ... + (c_n - 1) = k - 1, \tag{2.15}$$

since at each split time, the number of blocks increase by $c_i - 1$, and over the period $[0, T]$, the number of blocks increases by $k - 1$.

For times $t_1 < ... < t_n \in [0, T]$, write

$$\{\pi : \eta_0 \prec^{dt_1} \eta_1 \prec^{dt_2} ... \prec^{dt_n} \eta_n\} = \bigcap_{i=1}^{n} \{\pi_{t_i} = \eta_{i-1}, \pi_{t_i + dt_i} = \eta_i\} \tag{2.16}$$

for the event that for each $i = 1, ..., n$, the value of (π_t^T) changes from η_{i-1} to η_i in the time interval $(t_i, t_i + dt_i]$.

The following theorem gives the density of the event (2.16), thereby providing an alternate characterisation of the law of (π_t^T).

Theorem 2.2.3 (Split time representation for the law of (π_t^T))**.**

$$\mathbb{P}(\pi^{k,L,T} : \eta_0 \to^{dt_1} \eta_1 \to^{dt_2} ... \to^{dt_n} \eta_n) = \int_0^1 \frac{(1-s)^{k-1} F_T^1(s)}{(k-1)!\mathbb{P}(N_T \geq k)} \prod_{i=1}^n \Big(F_{T-t_i}^1(s)^{c_i - 1} f^{c_i}(F_{T-t_i}(s)) dt_i \Big) ds. \tag{2.17}$$

Note that for an integer j, $\mathbb{P}(L \geq j) = 0$ implies the j^{th}-derivative of the offspring generating function $f^j = 0$, and hence almost surely we have $c_i < j$ for every i.

We end this section by shedding light on the combinatorial structure of (2.14), by making reference to Faà Di Bruno's formula [23]. This formula states that for k-times differentiable functions f and g,

$$(\frac{d}{dx})^k (f \circ g)(x) = \sum_{\gamma \in \mathcal{P}_k} f^{|\gamma|}(g(x)) \prod_{\Gamma \in \gamma} g^{|\Gamma|}(x)$$

where $\Gamma \in \gamma$ refers to the blocks Γ in the partition γ, $|\Gamma|$ is the size of a block, and $|\gamma|$ is the number of blocks in γ. By setting $f = f_0$, and $g = f_1 \circ ... \circ f_n$, we can use induction to generalise this result to an $(n+1)$-fold composition:

$$(\frac{d}{dx})^k (f_0 \circ f_1 \circ ... \circ f_n)(x) = \sum_{(\gamma_0, \gamma_1, ..., \gamma_{n+1}) \in \mathcal{CP}_k^n} \prod_{i=0}^n \prod_{\Gamma_{i,j} \in \gamma_i} (f_i^{b_{i,j}} \circ f_{i+1} \circ ... \circ f_n)(x) \tag{2.18}$$

where b_{ij} are the breakage numbers of the chain $\gamma_0, ..., \gamma_{n+1}$.

Now suppose we have a one-parameter collection of smooth transformations $\{G_t : [0,1] \to [0,1] : t \geq 0\}$ with semigroup structure

$$G_0 = \mathsf{Id}, \; G_u \circ G_v = G_{u+v}. \tag{2.19}$$

Then for any sequence $0 = t_0 < t_1 ... < t_n < t_{n+1} = T$, writing $\Delta t_i = t_{i+1} - t_i$, we have for every $i = 0, 1..,$

$$G_{T-t_i} = G_{\Delta t_i} \circ ... \circ G_{\Delta t_n}$$

and hence by setting $f_i = G_{\Delta t_i}$ in (2.18) we can characterise the s derivatives of G_T through the equation

$$G_T^k(s) = \sum_{(\gamma_0, \gamma_1, ..., \gamma_{n+1}) \in \mathcal{CP}_k^n} \prod_{i=0}^n \prod_{\Gamma_{i,j} \in \gamma_i} G_{\Delta t_i}^{b_{i,j}} \circ G_{T-t_{i+1}}(s). \tag{2.20}$$

By the branching identity, the collection of smooth functions $\{F_t : t \geq 0\}$ given by

$$F_t(s) = \mathbb{E}[s^{N_t}]$$

obey the semigroup structure in (2.19). Noting that $F_t^j(s) \geq 0$, for every (j,t,s), it can be seen that $K(\cdot\,|s)$, defined by the finite dimensional distributions

$$K(\pi_{t_i} = \gamma_i | s) = \frac{\prod_{i=0}^{n} \prod_{\Gamma_{i,j} \in \gamma_i} F_{\Delta t_i}^{b_{i,j}} \circ F_{T-t_{i+1}}(s)}{F_T^k(s)}, \tag{2.21}$$

forms a probability measure on the product space $\mathcal{P}_k^{[0,T]}$.

Furthermore, we prove in Lemma 2.4.2 that M given by

$$M(ds) = \frac{(1-s)^{k-1} F_T^k(s)}{(k-1)! \mathbb{P}(N_T \geq k)} ds, \;\; s \in [0,1], \tag{2.22}$$

is a probability measure. It follows that (2.14) may be rewritten as mixture

$$\mathbb{P}(\pi_{t_1} = \gamma_1, ..., \pi_{t_n} = \gamma_n) = \int_0^1 M(ds) K(\pi_{t_i} = \gamma_i, \; \forall i | s). \tag{2.23}$$

That is, (π_t^T) may be interpreted as a mixture (2.22) of stochastic processes with finite dimensional distributions given by the Faà di Bruno quotients (2.21).

Alternatively, the conditional measure $K(\cdot|s)$ can be understood through the split time formula

$$K(\pi : \eta_0 \prec^{dt_1} \eta_1 \prec^{dt_2} ... \prec^{dt_n} \eta_n | s) = \frac{F_T^1(s)}{F_T^k(s)} \prod_{i=1}^{n} \Big(F_{T-t_i}^1(s)^{c_i - 1} f^{c_i}(F_{T-t_i}(s)) dt_i \Big) ds. \tag{2.24}$$

2.2.2 Asymptotic results

We now move on to our main results concerning the asymptotic behaviour of the partition process as $T \to \infty$. Below, we say a collection of $\mathcal{P}_k$-valued processes $\{\lambda^T = (\lambda_t^T)_{t\in[0,T]} : T > 0\}$ converge in distribution to a limit process $\bar{\lambda} = (\bar{\lambda}_t)_{t\in[0,\infty)}$ as $T \to \infty$, and we write

$$(\lambda_t^T)_{t\in[0,T]} \to^D (\bar{\lambda}_t)_{t\in[0,\infty)}$$

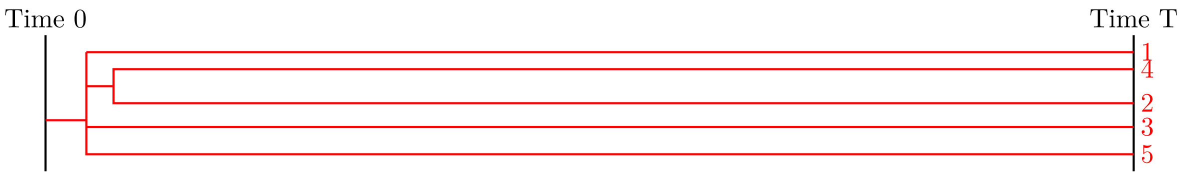
if for any fixed times $t_1 < .... < t_n$ and partition chain $\gamma_1 \prec ... \prec \gamma_n$ we have

$$\lim_{T\to\infty} \mathbb{P}(\lambda_{t_i}^T = \gamma_i \; \forall i) = \mathbb{P}(\bar{\lambda}_{t_i} = \gamma_i \; \forall i). \tag{2.25}$$

The finiteness of $\mathcal{P}_k$ ensures equation (2.25) is sufficient to ensure the weak convergence in distribution under the Skorokhod topology. See [14, Chapter III, Section 7.8] for details.

The Supercritical Case

We start with the supercritical regime $\mathbb{E}[L] > 1$, and further suppose that the Kesten-Stigum condition $\mathbb{E}[L \log L_+] < \infty$ holds. If we pick k individuals at a large time T from a supercritical tree, it turns out that the common ancestors of these k individuals last existed near the start of $[0,T]$.



In this case the martingale $N_t e^{-(m-1)t}$ converges to a non-degenerate limit W. See [5, Chapter III, Section 7] for details. In light of this almost-sure convergence, we may manipulate the limit

$$\lim_{T\to\infty} F(e^{-ve^{-(m-1)T}}, T) = \mathbb{E}[e^{-vW}] =: \varphi^0(v) \tag{2.26}$$

in the split time representation, equation (2.17), to prove the following.

Theorem 2.2.4. Suppose $\mathbb{E}[L] > 1$ and $\mathbb{E}[L \log_+ L] < \infty$. Then there exists a limit process $(\bar{\pi}_t^{k,L})_{t\in[0,\infty)}$ such that

$$(\pi_t^{k,L,T})_{t\in[0,T]} \to^D (\bar{\pi}_t^{k,L})_{t\in[0,\infty)}$$

and the limit process has law given by

$$\mathbb{P}(\bar{\pi}^{k,L} : \eta_0 \prec^{dt_1} \eta_1 \prec^{dt_2} \dots \prec^{dt_n} \eta_n) = \int_0^\infty \frac{v^{k-1}\varphi^1(v)}{(k-1)!(1-\varphi^0(\infty))} \prod_{i=1}^n g^{c_i}(t;v)dv \tag{2.27}$$

where $\varphi^j(v) = \mathbb{E}[W^j e^{-vW}]$, and the $g_{c_i}(\cdot|v)$ are non-unit densities given by

$$g_{c_i}(t|v) = \left(\varphi^1(ve^{-(m-1)t_i})e^{-(m-1)t_i}\right)^{c_i-1} f^{c_i}(\varphi^0(ve^{-(m-1)t_i}))dt_i.$$

This theorem is proved in section 4. The main idea of the proof is to change variable $s(v) = e^{-ve^{-r(m-1)T}}$ to plug (2.26) into (2.17).

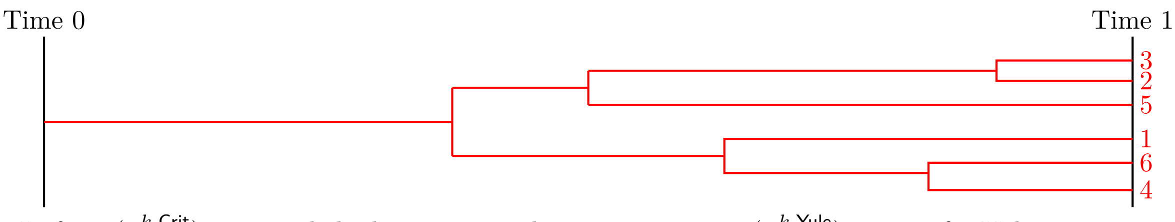
The Critical Case

Let us briefly mention here the following result from [20]. Suppose we are in the critical case with $\mathbb{E}[L^2] < \infty$. We find the ancestral partition process (π_t^T) gets scaled with the interval $[0,T]$ as $T \to \infty$.

Theorem 2.2.5. (Harris, Johnston and Roberts, 2017)
If the process is critical and $\mathbb{E}[L^2] < \infty$, then

$$(\pi_{tT}^{k,L,T})_{t\in[0,1]} \xrightarrow{D} (\bar{\pi}_t^{k,\mathsf{Crit}})_{t\in[0,1]}, \tag{2.28}$$

and the limit process $(\bar{\pi}_t^{k,\mathsf{Crit}})$ is universal in critical L with $\mathbb{E}[L^2] < \infty$. Furthermore, $(\bar{\pi}_t^{k,\mathsf{Crit}})$ is a binary tree topologically equivalent to a Kingman coalescent [28], but with more complicated merger times.



In fact, $(\bar{\pi}_t^{k,\mathsf{Crit}})_{t\in[0,1]}$ and the limit ancestral partition process $(\bar{\pi}_t^{k,\mathsf{Yule}})_{t\in[0,\infty)}$ of a Yule tree are deterministic time changes of one another, by the transformation

$$s \mapsto 1 - e^{-s}.$$

For further details on this fascinating link and on other properties of the process $(\bar{\pi}_t^{k,\mathsf{Crit}})$, see [20].

The Subcritical Case

As mentioned earlier, on the rare event that a subcritical process survives until time T, it is overwhelmingly likely that just one lineage did the surviving for the majority of $[0, T]$. Hence, we expect the most recent common ancestor of k individuals chosen at a large time T to be close to the time T.

For a subcritical process, the conditional limit law states that the conditional random variable $N_T | \{N_T \geq 1\}$ has a quasi-stationary limit. Namely, for each $j \in \{1, 2, \ldots\}$ the limit $a_j := \lim_{T\to\infty} \mathbb{P}(N_T = j | N_T \geq 1)$ exists, and that $\sum_{j=1}^{\infty} a_j = 1$. See [5, Chapter III, Section 7] for details. Write W for a random variable with distribution $\mathbb{P}(W = j) = a_j$, and $B(s) = \mathbb{E}[s^W] = \sum_{n\geq 1} a_n s^n$ for the associated quasi-stationary generating function.

Suppose we have a sequence of merging partitions $[\{1\}, ..., \{k\}] = \eta_0 \succ \eta_1 \succ ... \succ \eta_n = [\{1, 2, ..., k\}]$, with the property that η_i is obtained from η_{i-1} by c_i blocks of η_{i-1} merging to form precisely one block in η_i. For a coalescent process (ρ_t), write

$$\{\rho : \eta_0 \succ^{dt_1} \eta_1 \succ^{dt_2} ... \succ^{dt_n} \eta_n\} = \bigcap_{i=1}^{n} \{\rho_{t_i} = \eta_{i-1}, \rho_{t_i + dt_i} = \eta_i\} \tag{2.29}$$

Theorem 2.2.6. Suppose $\mathbb{E}[L] < 1$. Then there exists a limit process $(\bar{\rho}_t^{k,L})_{t\in[0,\infty)}$ such that

$$(\rho_t^{k,L,T})_{t\in[0,T]} \to^D (\bar{\rho}_t^{k,L})_{t\in[0,\infty)},$$

and the limit process has law given by merger times

$$\mathbb{P}(\bar{\rho}^{k,L} : \eta_0 \succ^{dt_1} \eta_1 \succ^{dt_2} ... \succ^{dt_n} \eta_n) = \int_0^1 \frac{(1-s)^{k-1} B'(s)}{(k-1)!\mathbb{P}(W \geq k)} \prod_{i=1}^{n} \Big\{ F'_{t_i}(s)^{c_i - 1} f^{c_i}(F_{t_i}(s)) dt_i \Big\} ds$$

We prove this in section 5. The main idea is to use the conditional limit law in (2.17).

2.3. Spines partitions and changes of measure

In this section we introduce spines, our tool for calculating the distributions of ancestral partition processes of uniformly chosen individuals. For each $n \in \mathbb{N}$, we associate a line of descent $(\xi_t^n)_{t\geq 0}$ that flows through the tree forward in time, choosing uniformly a branch to follow next at branching points. Our idea is to create a change of measure under which the first k spines $n = 1, ..., k$ flow through a tree forward in time in such a way that at time T, the particles carrying spines $1, ..., k$ are distinct, and represent a uniform sample of k individuals from the N_T alive.

During section 3 we will assume that for every k, $\mathbb{E}[L^k] < \infty$ and hence $\mathbb{E}[N_t^k] < \infty$ for all t. A result in section 5 shows we can safely remove this assumption.

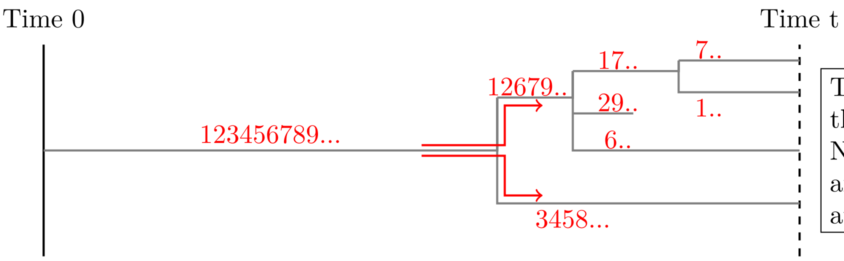
2.3.1 $\mathbb{N}$ spines

Suppose under a measure $\mathbb{P}$ we have a continuous-time Galton-Watson branching process where particles live for a exponential rate 1 amount of time, and upon death are replaced by a random number of partiticles distributed like L. Recall we write $\mathcal{N}_t$ for the set of individuals alive at time t, and $N_t = |\mathcal{N}_t|$. For technical reasons, we append a cemetery state Δ to the statespace, and write $\hat{\mathcal{N}}_t = \mathcal{N}_t \cup \Delta$.

Additionally under $\mathbb{P}$, for each $n \in \mathbb{N}$, there is a right-continuous stochastic process $(\xi_t^n)_{t\geq 0}$ called the n-spine defined as follows.

- At time t the n-spine takes values in $\hat{\mathcal{N}}_t$. That is, for some $u \in \hat{\mathcal{N}}_t$, $\xi_t^n = u$. We say that the n-spine is following u, and that u is carrying the n-spine.
- If a particle carrying the n-spine just before time t dies at time t and is replaced by $p \geq 1$ particles $v_1, ..., v_p$, then the n-spine chooses uniformly among the p offspring a particle to follow next. If the particle carrying the n-spine dies and is replaced by no offspring, we send the n-spine to the cemetery state Δ for the remainder of time. That is, $\xi_s^n = \Delta$ for all $s \geq t$.
- The n-spines don't affect the behaviour of the particles they are following. That is, if particle u is carrying the n-spine at time t, then this particle still branches at rate 1 and has offspring distributed like L.
- The set of n-spines $\{(\xi_t^n)_{t\geq 0} : n \in \mathbb{N}\}$ are independent of one another - that is, if a particle carrying some spines dies and is replaced by p offspring, each of these spines chooses uniformly an offspring to follow next, independently of the others.

So in essence, the n-spines are simply a set of labels that flow forward in time through a continuous-time Galton-Watson tree without affecting the law of the underlying tree.



The spines flow through the tree forward in time. Note that spines 2 and 9 are in the cemetery state at time t.

2.3.2 The spine partition change of measure $\mathbb{Q}^{\alpha,T}$

In order to avoid confusion about different measures, for the remainder of this section we write $\mathbb{P}[\cdot]$ rather than $\mathbb{E}[\cdot]$ for the expectation operator associated with $\mathbb{P}$.

For any set S and $k \geq 0$, let $S^{(k)}$ be the set of distinct k-tuples from S, and for $n \geq 0$, write

$$n^{(k)} = \begin{cases} n(n-1)(n-2)\dots(n-k+1) & \text{if } n \geq k \\ 0 & \text{otherwise.} \end{cases}$$

Note that $|S^{(k)}| = |S|^{(k)}$. Let $\mathcal{F}_t^{\varnothing}$ be the σ-algebra containing all the information about the tree up until time t, but without any knowledge of which particles the spines are following. For a subset $A \subseteq \mathbb{N}$, we call the set of processes $\{(\xi_t^A)_{t\geq 0} : a \in A\}$ the A-spines. Let $(\mathcal{F}_t^A)_{t\geq 0}$ be the filtration containing all the information about the entire tree and who the A-spines have been following until time t. Formally,

$$\mathcal{F}_t^A = \sigma\Big(\mathcal{F}_t^{\varnothing}; (\xi_s^a)_{s\in[0,t]}, a \in A\Big).$$

Of course if $A \subset B$, then $\mathcal{F}_t^A \subset \mathcal{F}_t^B$ for each t.

We now examine the probabilities, conditional on $\mathcal{F}_T^{\varnothing}$-knowledge, that a given spine is following a given particle in $\mathcal{N}_T$. For a particle in $u \in \mathcal{N}_T$, let

$$Q(u) = \prod_{v<u} L_v$$

be the product of birth sizes of ancestors of v.

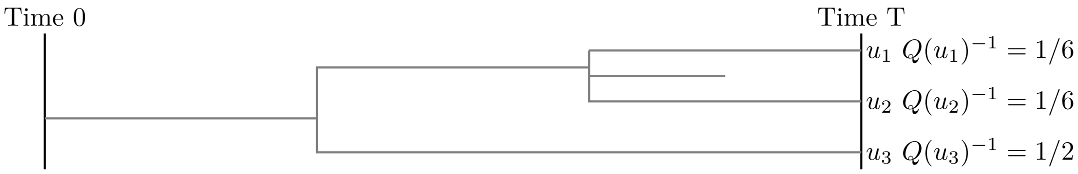
Note that for spine n to be following particle $u \in \mathcal{N}_T$, for each ancestor v of u, spine n must have chosen the 'correct' offspring of the L_v offspring of v to continue following. Hence,

$$\mathbb{P}(\xi_T^n = u|\mathcal{F}_T^{\varnothing}) = Q(u)^{-1} \tag{2.30}$$

for all $n \in \mathbb{N}$. Since the spines behave independently of one another, the probability that the A-spines are following a list $(u_a : a \in A)$ of (possibly non-distinct) members of $\mathcal{N}_t$ is

$$\mathbb{P}\big(\cap_{a\in A}\{\xi_t^a = u_a\}\big|\mathcal{F}_t^{\varnothing}\big) = \prod_{a\in A} Q(u_a)^{-1} \tag{2.31}$$

Since, in general, the quantities $Q(u)^{-1}$ in (2.30) vary for different $u \in \mathcal{N}_T$, under $\mathbb{P}$ the spines are more likely to be following some particles than others. This can be seen in the following example.



Now suppose $\alpha = [A_1, ..., A_k]$ is a partition of the finite set $A \subseteq \mathbb{N}$ into k blocks. For a fixed T, we will construct a change of measure $\mathbb{Q}^{\alpha,T}$ on $\mathcal{F}_T^A$ under which we have the following:

- There are always at least k particles alive at time T, and the A-spines are assigned to particles in $\mathcal{N}_T$ in such a way that the the following holds:

$$a \sim_\alpha b \iff \xi_T^a = \xi_T^b,\ \mathbb{Q}^{\alpha,T} - \text{almost surely.} \tag{2.32}$$

  That is, $a \sim_\alpha b$ if and only if spines a and b are following the same particle at time T.

- Given $\mathcal{F}_T^{\varnothing}$, the k-tuple of particles which the spines are following at time T is equally likely to be any k-tuple of those alive. That is, for a particular k-tuple $(u_1, ..., u_k) \in \mathcal{N}_T^{(k)}$,

$$\mathbb{Q}^{\alpha,T}\Big(\cap_{i=1}^k \cap_{a\in A_i}\{\xi_t^a = u_i\}\Big|\mathcal{F}_T^{\varnothing}\Big) = \frac{1}{N_T^{(k)}}.$$

We are ultimately interested in the case where $A = \{1, ..., k\}$ and $\alpha = [\{1\}, ..., \{k\}]$, because in this case, this change of measure forces the spines to represent a uniform sample of k particles from a population at time T.

For $A \subseteq \mathbb{N}$, let π_t^A be the partition of A defined by the equivalence relation

$$a \sim b \iff \xi_t^a = \xi_t^b,$$

that is, a and b are related in π_t^A if the a-spine and the b-spine are following the same particle at time t. Then (2.32) reads

$$\mathbb{Q}^{\alpha,T}(\pi_T^A = \alpha) = 1.$$

We define, for $t \geq 0$, the $\mathcal{F}_t^A$ measurable random variables

$$g_{\alpha,t} = \mathbb{1}\{\pi_t^A = \alpha\} \prod_{a \in A} Q(\xi_T^a) \tag{2.33}$$

Lemma 2.3.1. For any $t \geq 0$, $\mathbb{P}[g_{\alpha,t} | \mathcal{F}_t^{\varnothing}] = N_t^{(k)}$.

Proof. Note that every combination of k distinct particles $u = (u_1, ..., u_k) \in \mathcal{N}_t^{(k)}$ corresponds to a way in which $\pi_t^A = \alpha$, by setting $\xi_t^a = u_i, \forall a \in A_i, \forall\ i = 1, ..., k$. Thus

$$\begin{aligned}
\mathbb{P}[g_{\alpha,t} | \mathcal{F}_t^{\varnothing}] &= \mathbb{P}\Big[\mathbb{1}\{\pi_t^A = \alpha\} \prod_{a \in A} Q(\xi_T^a) \Big| \mathcal{F}_t^{\varnothing}\Big] \\
&= \mathbb{P}\Big[\sum_{u \in \mathcal{N}_t^{(k)}} \mathbb{1}\{\xi_t^a = u_i\ \forall a \in A_i,\ \forall\ i = 1, .., k\} \mathbb{1}\{\pi_t^A = \alpha\} \prod_{a \in A} Q(\xi_T^a) \Big| \mathcal{F}_t^{\varnothing}\Big] \\
&= \mathbb{P}\Big[\sum_{u \in \mathcal{N}_t^{(k)}} \prod_{i=1}^k \prod_{a \in A_i} \mathbb{1}\{\xi_t^a = u_i\} Q(\xi_T^a) \Big| \mathcal{F}_t^{\varnothing}\Big] \\
&= \sum_{u \in \mathcal{N}_t^{(k)}} \mathbb{P}\Big[\prod_{i=1}^k \prod_{a \in A_i} \mathbb{1}\{\xi_t^a = u_i\} Q(u_i) \Big| \mathcal{F}_t^{\varnothing}\Big]
\end{aligned}$$

since $\mathcal{N}_t \in \mathcal{F}_t^{\varnothing}$. Now since the spines are independent,

$$\mathbb{P}\Big[\prod_{i=1}^k \prod_{a \in A_i} \mathbb{1}\{\xi_t^a = u_i\} Q(u_i) \Big| \mathcal{F}_t^{\varnothing}\Big] = \prod_{i=1}^k \prod_{a \in A_i} \mathbb{P}\left[\mathbb{1}\{\xi_t^a = u_i\} Q(u_i) \Big| \mathcal{F}_t^{\varnothing}\right].$$

Finally, $Q(u_i) \in \mathcal{F}_t^{\varnothing}$, so for every a, i,

$$\mathbb{P}\left[\mathbb{1}\{\xi_t^a = u_i\} Q(u_i) \Big| \mathcal{F}_t^{\varnothing}\right] = Q(u_i) \mathbb{P}(\xi_t^a = u_i | \mathcal{F}_t^{\varnothing}) = 1$$

by (2.30), and hence

$$\mathbb{P}\Big[\prod_{i=1}^k \prod_{a \in A_i} \mathbb{1}\{\xi_t^a = u_i\} Q(u_i) \Big| \mathcal{F}_t^{\varnothing}\Big] = \prod_{i=1}^k \prod_{a \in A_i} 1 = 1. \tag{2.34}$$

Finally,

$$\mathbb{P}[g_{\alpha,t}|\mathcal{F}_t^{\varnothing}] = \sum_{u\in\mathcal{N}_t^{(k)}} 1 = |\mathcal{N}_t^{(k)}| = N_t^{(k)}.$$

$\square$

Let $A \subseteq \mathbb{N}$ be finite, α be a partition of A into k blocks, and T be a fixed time. By the previous lemma $\mathbb{P}[g_{\alpha,T}] = \mathbb{P}[\mathbb{P}[g_{\alpha,T}|\mathcal{F}_T^{\varnothing}]] = \mathbb{P}[N_T^{(k)}]$, so the random variable

$$\zeta_{\alpha,T} = \frac{g_{\alpha,T}}{\mathbb{P}[N_T^{(k)}]}$$

has unit mean.

Define a new probability measure $\mathbb{Q}^{\alpha,T}$ on $\mathcal{F}_T^A$ by setting

$$\left.\frac{\mathrm{d}\mathbb{Q}^{\alpha,T}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_T^A} = \zeta_{\alpha,T}. \tag{2.35}$$

Let

$$Z_{k,T} = \frac{N_T^{(k)}}{\mathbb{P}[N_T^{(k)}]}$$

so that, again by Lemma 2.3.1,

$$\left.\frac{\mathrm{d}\mathbb{Q}^{\alpha,T}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_T^{\varnothing}} = Z_{k,T}. \tag{2.36}$$

The remainder of section 3 is dedicated to studying the properties of $\mathbb{Q}^{\alpha,T}$, and how it affects the behaviour of the A-spines. In section 2.3.3, we show that the measure $\mathbb{Q}^{\alpha,T}$ has a nice uniformity property, namely that at time T, the A-spines are equally likely to be any k-tuple of particles alive at a time T. In section 2.3.4, we see the change of measure $\mathbb{Q}^{\alpha,T}$ has a key symmetry property.

2.3.3 Uniformity properties of $\mathbb{Q}^{\alpha,T}$

Let α be a partition of A into k blocks. The goal of section 2.3.3 is to prove that under $\mathbb{Q}^{\alpha,T}$, conditional on $\mathcal{F}_T^{\varnothing}$, the particles the A-spines are following at the time T are equally likely to be any k-tuple alive.

Lemma 2.3.2. Suppose that μ and ν are probability measures on the σ-algebra $\mathcal{F}$, and that $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$. If

$$\left.\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right|_{\mathcal{F}} = Y \quad \text{and} \quad \left.\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right|_{\mathcal{G}} = Z,$$

then for any non-negative $\mathcal{F}$-measurable X,

$$Z\mu[X|\mathcal{G}] = \nu[XY|\mathcal{G}] \quad \nu\text{-almost surely.}$$

Proof. For any $A \in \mathcal{G}$,

$$\nu[XY\mathbb{1}_A] = \mu[X\mathbb{1}_A] = \mu[\nu[X|\mathcal{G}]\mathbb{1}_A] = \nu[Z\nu[X|\mathcal{G}]\mathbb{1}_A].$$

Since $Z\mu[X|\mathcal{G}]$ is $\mathcal{G}$-measurable, it therefore satisfies the definition of conditional expectation of XY with respect to $\mathcal{G}$ under ν. $\square$

Applying this result using (2.35) and (2.36), we find that for any non-negative $\mathcal{F}_T^A$-measurable random variable X, on the event $\{Z_{k,T} > 0\}$,

$$\mathbb{Q}^{\alpha,T}[X|\mathcal{F}_T^{\varnothing}] = \frac{1}{Z_{k,T}}\mathbb{P}[X\zeta_{\alpha,T}|\mathcal{F}_T^{\varnothing}], \tag{2.37}$$

For a subinterval $I \subseteq [0,T]$, let $\mathcal{G}_I^A$ be the σ-algebra with knowledge of who the A-spines have been following during I, and the offspring sizes of ancestors of the spines during this interval, that is

$$\mathcal{G}_I^A = \sigma\Bigg(\Big\{\{\xi_r^a = u\} : u \in \mathcal{N}_s,\ s \in I, a \in A\Big\} \cup \Big\{L_v : \xi_s^a \leq v < \xi_t^a, s < t \in I, a \in A\Big\}\Bigg) \tag{2.38}$$

and set $\mathcal{G}_t^A = \mathcal{G}_{[0,t]}^A$.

Using (2.35) and the fact that $\zeta_{\alpha,T}$ is $\mathcal{G}_T^A$ measurable, we deduce that

$$\frac{\mathrm{d}\mathbb{Q}^{\alpha,T}}{\mathrm{d}\mathbb{P}}\bigg|_{\mathcal{G}_T^A} = \zeta_{\alpha,T}. \tag{2.39}$$

Applying Lemma 3.4.3, this time using (2.35) and (2.39), we deduce that for any $\mathcal{F}_T^A$ measurable random variable X,

$$\mathbb{Q}^{k,T}[X|\mathcal{G}_T^A] = \frac{1}{\zeta_{\alpha,T}}\mathbb{P}[X\zeta_{\alpha,T}|\mathcal{G}_T^A] = \mathbb{P}[X|\mathcal{G}_T^A]. \tag{2.40}$$

on the event $\{\zeta_{\alpha,T} > 0\}$; in the second equation we have used the $\mathcal{G}_T^A$-measurability of $\zeta_{\alpha,T}$.

In particular, (2.40) implies that $\mathbb{Q}^{\alpha,T}$ and $\mathbb{P}$ coincide on the collection of $\mathcal{F}_T^A$-measurable events that are independent of $\mathcal{G}_T^A$. In other words, particles not carrying any of the A-spines behave under $\mathbb{Q}^{\alpha,T}$ exactly as they do under $\mathbb{P}$: they branch at unit rate and have offspring distribution L.

Note by the definition of $\zeta_{\alpha,T}$ and Lemma 2.3.1 that

$$\mathbb{Q}^{\alpha,T}(\pi_T^A = \alpha) = 1.$$

Now if $|\alpha| = k$, there must be at least k distinct particles alive at time T for the spines to follow. The previous display therefore yields

$$\mathbb{Q}^{\alpha,T}(N_T \geq k) = 1.$$

In summary, under $\mathbb{Q}^{\alpha,T}$ the A-spines at time T are distributed across k different particles in $\mathcal{N}_T$ and induce the partition α of A at time T. The following lemma tells us that these k particles are chosen uniformly without replacement from those alive at time T.

Lemma 2.3.3. When $|\alpha| = k$, the $\mathbb{Q}^{\alpha,T}$-probability that the spines are following a particular k-tuple $(u_1, ..., u_k) \in \mathcal{N}_T^{(k)}$ equals $1/N_T^{(k)}$. That is

$$\mathbb{Q}^{\alpha,T}\Big(\cap_{i=1}^k \cap_{a\in\alpha_i}\{\xi_t^a = u_i\}\Big|\mathcal{F}_T^{\varnothing}\Big) = \frac{1}{N_T^{(k)}}.$$

Proof. Note that if $N_T \geq k$ then $Z_{k,T} > 0$. Then by (2.37), for any $u \in \mathcal{N}_T^{(k)}$,

$$\begin{aligned}
\mathbb{Q}\bigg(\cap_{i=1}^k \cap_{a\in\alpha_i} \{\xi_t^a = u_i\} \bigg| \mathcal{F}_T^{\varnothing} \bigg) &= \frac{1}{Z_{k,T}} \mathbb{P}\bigg[\zeta_{\alpha,T} \mathbb{1}\{\cap_{i=1}^k \cap_{a\in\alpha_i} \{\xi_t^a = u_i\}\} \bigg| \mathcal{F}_T^{\varnothing} \bigg] \\
&= \frac{\mathbb{P}[N_T^{(k)}]}{N_T^{(k)}} \frac{1}{\mathbb{P}[N_T^{(k)}]} \mathbb{P}\bigg[\prod_{i=1}^k \prod_{a\in\alpha_i} \mathbb{1}\{\xi_t^a = u_i\} Q(u_i) \bigg| \mathcal{F}_t^{\varnothing} \bigg] \\
&= \frac{1}{N_T^{(k)}}
\end{aligned}$$

where the third equality follows from (2.34). $\square$

2.3.4 The law of $(\pi_t^A)_{t\in[0,T]}$ under $\mathbb{Q}^{\alpha,T}$ and symmetric properties

Let α be a partition of a finite subset $A \subseteq \mathbb{N}$, and let $[A] = \gamma_0 \prec ... \prec \gamma_n \prec \gamma_{n+1} = \alpha$ be a chain of partitions.

As we have seen, under $\mathbb{Q}^{\alpha,T}$ the partition process $(\pi_t^A)_{t\in[0,T]}$ starts as the one block partition $[A]$ at time 0 and finishes at the partition α at time T. In this section, we will prove a theorem that describes the finite dimensional distributions of this partition process on $[0,T]$ under the change of measure $\mathbb{Q}^{\alpha,T}$. Furthermore, we give an essential symmetry property of the measure $\mathbb{Q}^{\alpha,T}$, which is the crux of the entire paper.

In order to understand this symmetry property, recall from the above that $\mathcal{G}_{[s,t]}^B$ is the σ-algebra with all the information about the B-spines during the time period $[s,t]$. Define the shift operator

$$\theta_s : \mathcal{G}_{[s,t]}^B \to \mathcal{G}_{[0,t-s]}^B$$

informally as the map that takes an event occuring during $[s,t]$ and maps it to the corresponding event occuring during $[0, t-s]$. For example, for $0 < r < t-s$, and for $b, b' \in B$,

$$\theta_s\Big(\big\{\text{Spines } b \text{ and } b' \text{ together at time } s+r\big\}\Big) = \big\{\text{Spines } b \text{ and } b' \text{ together at time } r\big\}.$$

The symmetry property states the following. Let B be a subset of A, and let $(\pi_t^B)_{t\in[0,T]}$ be the partition induced by the B-spines. For $s < t$, conditional on the event

$$\{\pi_s^B = [B], \pi_t^B = \beta\}$$

the behaviour of the B-spines under $\mathbb{Q}^{\alpha,T}$ during the time period $[s,t]$ looks like $\mathbb{Q}^{\beta,t-s}$.

In terms of shift operators, this says that for any event $W \in \mathcal{F}_{(s,t)}^B$

$$\mathbb{Q}^{\alpha,T}\Big(W \Big| \pi_s^B = [B], \pi_t^B = \beta\Big) = \mathbb{Q}^{\beta,t-s}\Big(\theta_s(W)\Big).$$

Theorem 2.3.4. Let $[A] = \gamma_0 \prec \gamma_1 \prec .. \prec \gamma_n \prec \gamma_{n+1} = \alpha$ be a partition chain, let $t_0 = 0 < t_1 < ... < t_n < t_{n+1} = T$, and $\Delta t_i = t_{i+1} - t_i$. Then, under $\mathbb{Q}^{\alpha,T}$,

1.

$$\mathbb{Q}^{\alpha,T}(\pi_{t_i}^A = \gamma_i \ \forall\ i) = \frac{\prod_{i=0}^n \prod_{\Gamma_{i,j}\in\gamma_i} \mathbb{P}[N_{\Delta t_i}^{(b_{ij})}]}{\mathbb{P}[N_T^{(k)}]} \tag{2.41}$$

2. Conditioned on the event $\{\pi^A_{t_i} = \gamma_i \ \forall\ i\}$, during a time interval (t_i, t_{i+1}), the spines in $\Gamma_{i,j}$ behave as if under $\mathbb{Q}^{\gamma_{ij},\Delta t_i}$, and the σ-algebras $\left\{\mathcal{G}^{\Gamma_{i,j}}_{[t_i,t_{i+1}]} : i = 0, 1, ..., n, \Gamma_{i,j} \in \gamma_i\right\}$ are conditionally independent.

Proof. We prove the result by induction on n. First, let the prove the result when $n = 1$. Let $t \in [0, T]$, and let $\gamma = [\Gamma_1, ..., \Gamma_p]$ be a partition such that $[A] \prec \gamma \prec \alpha$. Write γ_j for the restriction of the partition α to the block Γ_j.

Now suppose that $X \in \mathcal{G}^A_{[0,t]}$ and $Y_j \in \mathcal{G}^{\Gamma_j}_{[t,T]}, j = 1, ...p$ are events. We show that

$$\mathbb{Q}^{\alpha,T}\left(X \cap \bigcap_{j=1}^{p} Y_j, \pi_t = \gamma\right) = \frac{\mathbb{P}\left[N^{(p)}_t\right] \prod_{j=1}^{p} \mathbb{P}\left[N^{(k_j)}_{T-t}\right]}{\mathbb{P}[N^{(k)}_T]} \mathbb{Q}^{\gamma,t}(X) \prod_{j=1}^{p} \mathbb{Q}^{\gamma_j,T-t}(\theta_t(Y_j)) \tag{2.42}$$

To see that this implies the result for $n = 1$, the first part follows by plugging in $\mathbb{1}_X = 1, \mathbb{1}_{Y_j} = 1,\ \forall\ j$. Dividing through by (2.41) obtains the second part.

To prove (2.42), use (**??**) to write

$$\mathbb{Q}^{\alpha,T}\left(X \cap \bigcap_{j=1}^{p} Y_j, \pi_t = \gamma\right) = \frac{1}{\mathbb{P}[N^{(k)}_T]} \mathbb{P}\left[\mathbb{1}\{\pi^A_t = \gamma, \pi^A_T = \alpha\} \prod_{a\in A} Q(\xi^a_T)\ \mathbb{1}_X \prod_{j=1}^{p} \mathbb{1}_{Y_j}\right]$$

Note first

$$\mathbb{1}\{\pi^A_t = \gamma, \pi^A_T = \alpha\} = \mathbb{1}\{\pi^A_t = \gamma\} \prod_{j=1}^{p} \mathbb{1}\{\pi^{\Gamma_j}_T = \gamma_j\},$$

and that

$$\prod_{a\in A} Q(\xi^a_T) = \prod_{a\in A} Q(\xi^a_t) \prod_{j=1}^{p} \prod_{a\in\Gamma_j} \frac{Q(\xi^a_T)}{Q(\xi^a_t)}$$

Hence,

$$\mathbb{Q}^{\alpha,T}\left(X \cap \bigcap_{j=1}^{p} Y_j, \pi_t = \gamma\right) = \frac{\mathbb{P}\left[N^{(p)}_t\right] \prod_{j=1}^{p} \mathbb{P}\left[N^{(k_j)}_{T-t}\right]}{\mathbb{P}[N^{(k)}_T]} \mathbb{P}\left[H_0 \prod_{j=1}^{p} H_j\right]$$

where

$$H_0 = \frac{\mathbb{1}\{\pi^A_t = \gamma\} \prod_{a\in A} Q(\xi^a_t)}{\mathbb{P}[N^{(k)}_t]} \mathbb{1}_X$$

and

$$H_j = \frac{\mathbb{1}\{\pi^{\Gamma_j}_T = \gamma_j\} \prod_{a\in\Gamma_j} \frac{Q(\xi^a_T)}{Q(\xi^a_t)}}{\mathbb{P}[N^{k_j}_{T-t}]} \mathbb{1}_{Y_j}$$

Now here is the key observation. By the definition (2.39), $H_0 = \zeta_{\gamma,t}\mathbb{1}_X$, and hence

$$\mathbb{P}[H_0] = \mathbb{Q}^{\gamma,t}(X).$$

Furthermore, $\{H_0 > 0\}$ implies $\{\pi_t^A = \gamma\}$. Note that for each j H_j only depends on H_0 through the event $\{\pi_t^A = \gamma\}$, which implies $\{\pi_t^{\Gamma_j} = [\Gamma_j]\}$.

Now given that the Γ_j spines are following the same particle $u \in \mathcal{N}_t$ at time t, by the Markovian nature of the branching process, the subtree generated by descendents of u looks like a tree started at time 0. Furthermore, the random variable $Q(\xi_T^a)/Q(\xi_t^a)$ is a product of births experienced by the a-spine during the time period $(t, T]$, that is, like a copy of $Q(\tilde{\xi}^a_{T-t})$. It follows that conditionally on $\{\pi_t^{\Gamma_j} = [\Gamma_j]\}$, we have the following equivalence in law under $\mathbb{P}$:

$$\left.\frac{\mathbb{1}\{\pi_T^{\Gamma_j} = \gamma_j\} \prod_{a \in \Gamma_j} \frac{Q(\xi_T^a)}{Q(\xi_t^a)}}{\mathbb{P}[N_{T-t}^{k_j}]}\right| \{\pi_t^{\Gamma_j} = [\Gamma_j]\} \ \stackrel{\mathsf{Law}}{\equiv} \ \zeta_{\gamma_j, T-t},$$

which implies

$$\mathbb{P}[H_j | H_0 > 0] = \mathbb{P}[H_j | \pi_T^{\Gamma_j} = \gamma_j] = \mathbb{Q}^{\gamma_j, T-t}(\theta_t(Y_j))$$

Finally, on the event $\{\pi_t = \gamma\}$ clearly the collection of spine groups $\Gamma_j, j = 1, 2, ..., p$ are independent of one another during $[t, T]$ and the A-spines during $[0, t]$, hence

$$\mathbb{P}\left[H_0 \prod_{j=1}^p H_j\right] = \mathbb{Q}^{\gamma, t}(X) \prod_{j=1}^p \mathbb{Q}^{\gamma_j, T-t}(\theta_t(Y_j)),$$

and this establishes the result for $n = 1$.

It is straightforward to see the result holds for $n > 1$ by induction. Suppose the result holds for all $m = 1, ..., n$. Let $A \prec \gamma_1 \prec ... \gamma_n \prec \hat{\gamma} \prec \alpha$ be a partition chain of internal length $n+1$, and let $t_1 < ... < t_n < \hat{t} \in [0, T]$. Let the sets in $\hat{\gamma}$ have sizes $k_1 + ... k_p = k$. Then by the case $n = 1$,

$$\begin{aligned}
\mathbb{Q}^{\alpha, T}\left[\pi_{t_i}^A = \gamma_i \ \forall i = 1, ..., n, \ \pi_{\hat{t}}^A = \hat{\gamma}\right] &= \mathbb{Q}^{\alpha, T}(\pi_{\hat{t}} = \hat{\gamma}) \mathbb{Q}^{\hat{\gamma}, t}(\pi_{t_i}^A = \gamma_i \ \forall i = 1, ..., n) \\
&= \frac{\mathbb{P}[N_t^{(p)}] \prod_{j=1}^p \mathbb{P}[N_{T-t}^{(k_j)}]}{\mathbb{P}[N_T^{(k)}]} \frac{\prod_{i=0}^n \prod_{\Gamma_{i,j} \in \gamma_i} \mathbb{P}[N_{\Delta t_i}^{(b_{ij})}]}{\mathbb{P}[N_t^{(p)}]} \\
&= \frac{\prod_{i=0}^n \prod_{\Gamma_{i,j} \in \gamma_i} \mathbb{P}[N_{\Delta t_i}^{(b_{ij})}] \prod_{j=1}^p \mathbb{P}[N_{T-t}^{(k_j)}]}{\mathbb{P}[N_T^{(k)}]} \\
&= \frac{\prod_{i=0}^{n+1} \prod_{\Gamma_{i,j} \in \gamma_i} \mathbb{P}[N_{\Delta t_i}^{(b_{ij})}]}{\mathbb{P}[N_T^{(k)}]}
\end{aligned}$$

where we note that the terms $\prod_{j=1}^p \mathbb{P}[N_{T-t}^{(k_j)}]$ form the $(n+1)^{\text{th}}$ breakage numbers, so in the final line equality above we have called $\hat{\gamma} = \gamma_{n+1}$ and written $\Delta t_{n+1} = T - \hat{t}$.

The conditional independence follows immediately by induction and the case $n = 1$. On the event $\{\pi_{t_0} = [A], \pi_{t_1} = \gamma_1, ..., \pi_{t_n} = \gamma_n\}$, by the inductive hypothesis the σ-algebras $\left\{\mathcal{G}_{[t_i, t_{i+1}]}^{\Gamma_{i,j}} : i = 0, 1, ..., n, \Gamma_{i,j} \in \gamma_i\right\}$ are conditionally independent. The case $n = 1$ now deals with splitting the interval $[t_n, T]$ into $[t_n, \hat{t}]$ and $[\hat{t}, T]$. $\square$

2.3.5 Joint generating functions N_T and $(\pi_t^k)_{t\in[0,T]}$ under $\mathbb{Q}^{k,T}$

If α is the partition of $\{1,...,k\}$ into singletons, we write abuse notation and write $\mathbb{Q}^{k,T} := \mathbb{Q}^{\alpha,T}$. Similarly, we write k-spines rather than $\{1,...,k\}$-spines and π_t^k for $\pi_t^{\{1,...,k\}}$.

Our goal now is to use the symmetry theorem to calculate the joint generating function

$$\mathbb{Q}^{k,T}\left(s^{N_T-k}; \pi_{t_i}^k = \gamma_i \ \forall\ i\right)$$

Theorem 2.3.5. Let $F_t(s) = \mathbb{P}[s^{N_t}]$ be the generating function of the branching process, $0 = t_0 < t_1 < ... < t_n < t_{n+1} = T$, $\Delta t_i = t_{i+1} - t_i$ and let $\{1,...,k\} = \gamma_0 \prec \gamma_1 \prec ... \prec \gamma_n \prec \gamma_{n+1} = [\{1\},...,\{k\}]$ be a partition chain with breakage numbers b_{ij}. Then

$$\mathbb{Q}^{k,T}\left[s^{N_T-k}\ \mathbb{1}\left\{\pi_{t_i}^k = \gamma_i \ \forall\ i\right\}\right] = \frac{1}{\mathbb{P}[N_T^{(k)}]}\prod_{i=0}^{i=n}\prod_{\Gamma_{i,j}\in\gamma_i} F_{\Delta t_i}^{b_{ij}}(F_{T-t_{i+1}}(s))$$

Proof. Set $B = \{\pi_{t_i}^k = \gamma_i \ \forall\ i\}$. First we calculate the conditional generating function

$$\mathbb{Q}^{k,T}\left(s^{N_T-k}\middle| B\right).$$

At time T, there are N_T particles alive, k of whom are spines. Thus $N_T - k$ is equal to the number of individuals at time T not carrying a k-spine. We call these the non-carriers. For a non-carrier alive at time T, let

$$r_u := \sup\{r\in[0,T] : \exists l = 1,...,k,\ \xi_r^l < u\}$$

be the time to most recent spine ancestor of u. On the event B, for each non-carrier $u\in\mathcal{N}_T$, there is a unique time interval $[t_i, t_{i+1})$ containing r_u, and a unique $j = 1,...,|\gamma_i|$ such that a Γ_{ij}-spine comes arbitrarily close to this supremum. That is, a Γ_{ij} spine was the most recent spine ancestor of u during the time interval $[t_i, t_{i+1})$. With this in mind, on the event B, let $\mathcal{M}_T^{\Gamma_{i,j}}$ be the set of non-carriers whose most recent k-spine ancestor of u was $\Gamma_{i,j}$-spine during the interval $[t_i, t_{i+1})$. It follows that we may write

$$N_T - k = \sum_{i=0}^{n}\sum_{\Gamma_{i,j}\in\gamma_i} M_T^{\Gamma_{i,j}}$$

where $M_T^{\Gamma_{i,j}} = |\mathcal{M}_T^{\Gamma_{i,j}}|$. By Theorem 2.3.4, the behaviour of the spines in $\Gamma_{i,j}$ during the period $[t_i, t_{i+1})$ looks like $\mathbb{Q}^{\gamma_{ij},\Delta t_i}$, independently of other (i,j), and hence the terms of the sum are conditionally independent on B.

Furthermore, again conditional on B, the number of individuals at time t_{i+1} who were born off a spine in $\Gamma_{i,j}$ during the period $[t_i, t_{i+1})$ is distributed like $N_{\Delta t_i} - b_{ij}$ under $\mathbb{Q}^{\gamma_{ij},\Delta t_i}$, where we recall $b_{ij} = |\gamma_{ij}|$. Each one of these time t_{i+1} individuals goes on to grow as an independent population governed by $\mathbb{P}$ over the time period $[t_{i+1}, T]$. Write $N_{T-t_{i+1}}^{[i]}$ for these independent populations. Using the fact that $N_{T-t}^{[i]}$ are distributed like independent copies of N_{T-t} in the second equality below, the conditional generating function for $M_T^{\Gamma_{i,j}}$ is

$$\begin{aligned}
\mathbb{Q}^{k,T}\left(s^{M_T^{\Gamma_{ij}}}\big|B\right) &= \mathbb{Q}^{b_{ij},\Delta t_i}\left[s^{\left(\sum_{i=1}^{(N_{\Delta t_i}-b_{ij})} N^{[i]}_{T-t_{i+1}}\right)}\right] \\
&= \mathbb{Q}^{b_{ij},\Delta t_i}\left[\mathbb{P}[s^{N_{T-t_{i+1}}}]^{(N_{\Delta t_i}-b_{ij})}\right] \\
&= \mathbb{Q}^{b_{ij},\Delta t_i}\left(F_{T-t_{i+1}}(s)^{(N_{\Delta t_i}-b_{ij})}\right),
\end{aligned}$$

and using the conditional independence of the $M_T^{\Gamma_{ij}}$ we have

$$\begin{aligned}
\mathbb{Q}^{k,T}\left(s^{N_T-k}\big|B\right) &= \mathbb{Q}^{k,T}\left(s^{\sum_{i=0}^{n}\sum_{\Gamma_{i,j}\in\gamma_i} M_T^{\Gamma_{i,j}}}\big|B\right) \\
&= \prod_{i=0}^{i=n}\prod_{\Gamma_{i,j}\in\gamma_i}\mathbb{Q}^{b_{ij},\Delta t_i}\left(F_{T-t_{i+1}}(s)^{(N_{\Delta t_i}-b_{ij})}\right).
\end{aligned}$$

For some γ, t with $|\gamma| = k$ consider the quantity

$$\mathbb{Q}^{\gamma,t}[v^{N_t-k}].$$

Since $N_t - k$ is $\mathcal{F}_t^{\varnothing}$-measurable, (2.36) implies that

$$\mathbb{Q}^{\gamma,t}[v^{N_t-k}] = \mathbb{P}\Big[Z_{k,t}\ v^{N_t-k}\Big] = \mathbb{P}\left[\frac{N_t^{(k)}}{\mathbb{P}[N_t^{(k)}]}v^{N_t-k}\right] = \frac{1}{\mathbb{P}[N_t^{(k)}]}F_t^k(v)$$

In particular, with $\gamma = \gamma_{ij}$, $k = b_{ij} = |\gamma_{ij}|$, $t = \Delta t_i$ and $v = F_{T-t_{i+1}}(s)$, we have

$$\mathbb{Q}^{b_{ij},\Delta t_i}\left(F_{T-t_{i+1}}(s)^{(N_{\Delta t_i}-b_{ij})}\right) = \frac{F_{\Delta t_i}^{b_{ij}}\left(F_{T-t_{i+1}}(s)\right)}{\mathbb{P}[N_{\Delta t_i}^{(b_{ij})}]}$$

and thus

$$\mathbb{Q}^{k,T}\left(s^{N_T-k}\big|B\right) = \prod_{i=0}^{i=n}\prod_{\Gamma_{i,j}\in\gamma_i}\frac{F_{\Delta t_i}^{b_{ij}}(F_{T-t_{i+1}}(s))}{\mathbb{P}[N_{\Delta t_i}^{(b_{ij})}]}$$

Multiplying this equation by (2.41), we obtain

$$\mathbb{Q}^{k,T}\left[s^{N_T-k}\ \mathbb{1}\left\{\pi_{t_i}^k = \gamma_i\ \forall\ i\right\}\right] = \frac{1}{\mathbb{P}[N_T^{(k)}]}\prod_{i=0}^{i=n}\prod_{\Gamma_{i,j}\in\gamma_i}F_{\Delta t_i}^{b_{ij}}(F_{T-t_{i+1}}(s)) \tag{2.43}$$

□

2.4. Coalescent formula for uniformly chosen particles and proof of our main theorem

2.4.1 Expectation of functionals of uniformly chosen particles

Assume $\mathbb{P}[L^k] < \infty$. Suppose, conditional on $\{N_T \geq k\}$, we pick uniformly k distinct particles and label them $U_1, ..., U_k$, writing $\bar{U} = (U_1, ..., U_k)$. For a nonnegative $\mathcal{G}_T^k$-measurable functional

f, we want to calculate the expectation

$$\mathbb{P}[f(\bar{U})|N_T \geq k] = \mathbb{P}\left[\frac{1}{N_T^{(k)}} \sum_{\bar{u} \in \mathcal{N}_T^{(k)}} f(\bar{u}) \middle| N_T \geq k\right].$$

The following lemma permits us to give a value for this quantity in terms of the $\mathbb{Q}^{k,T}$-behaviour of the spines $\{1, ..., k\}$.

Lemma 2.4.1. Write $\bar{\xi}_t = (\xi_t^1, ..., \xi_t^k)$ for the vector of k-spines at time t.

$$\mathbb{P}[f(\bar{U})|N_T \geq k] = \frac{\mathbb{P}[N_T^{(k)}]}{\mathbb{P}(N_T \geq k)} \mathbb{Q}^{k,T}\Big[\frac{1}{N_T^{(k)}} f(\bar{\xi})\Big]$$

Proof. Note that

$$\begin{aligned}
\mathbb{Q}^{k,T}\Big[f(\bar{\xi})|\mathcal{F}_T^{\varnothing}\Big]\mathbb{1}_{N_T \geq k} &= \mathbb{Q}^{k,T}\Big[\sum_{\bar{u} \in \mathcal{N}_T^{(k)}} f(\bar{u})\mathbb{1}\{\bar{u} = \bar{\xi}\}\Big|\mathcal{F}_T^{\varnothing}\Big]\mathbb{1}_{N_T \geq k} \\
&= \sum_{\bar{u} \in \mathcal{N}_T^{(k)}} f(\bar{u})\mathbb{Q}^{k,T}\Big[\mathbb{1}\{\bar{u} = \bar{\xi}\}\Big|\mathcal{F}_T^{\varnothing}\Big]\mathbb{1}_{N_T \geq k} \\
&= \frac{\mathbb{1}_{N_T \geq k}}{N_T^{(k)}} \sum_{\bar{u} \in \mathcal{N}_T^{(k)}} f(\bar{u})
\end{aligned}$$

by Lemma 2.3.3. Hence

$$\begin{aligned}
\mathbb{P}\Big[\frac{1}{N_T^{(k)}} \sum_{\bar{u} \in \mathcal{N}_T^{(k)}} f(\bar{u})\Big|N_T \geq k\Big] &= \frac{1}{\mathbb{P}(N_T \geq k)}\mathbb{P}\Big[\mathbb{Q}^{k,T}\Big[f(\bar{\xi})|\mathcal{F}_T^{\varnothing}\Big]\mathbb{1}_{N_T \geq k}\Big] \\
&= \frac{1}{\mathbb{P}(N_T \geq k)}\mathbb{Q}^{k,T}\Big[\frac{1}{Z_{k,T}}\mathbb{Q}^{k,T}\Big[f(\bar{\xi})|\mathcal{F}_T^{\varnothing}\Big]\mathbb{1}_{N_T \geq k}\Big] \\
&= \frac{\mathbb{P}[N_T^{(k)}]}{\mathbb{P}(N_T \geq k)}\mathbb{Q}^{k,T}\Big[\frac{1}{N_T^{(k)}} f(\bar{\xi})\Big]
\end{aligned}$$

as required. □

Lemma 2.4.2. Let N be a $\{0, 1, ...\}$-valued random variable such that $P(N \geq k) > 0$. Then

$$\int_0^1 \frac{(1-s)^{k-1}\mathbb{E}[N^{(k)}s^{N-k}]}{(k-1)!\mathbb{P}(N \geq k)} ds = 1 \tag{2.44}$$

Proof. Recall the definition of the beta function:

$$B(x, y) = \frac{(x-1)!(y-1)!}{(x+y-1)!} = \int_0^1 s^{x-1}(1-s)^{y-1} ds.$$

It follows that for $n \geq k$, we have the identity

$$\frac{1}{n^{(k)}} = \frac{(n-k)!}{n!} = \frac{1}{(k-1)!}\int_0^1 (1-s)^{k-1}s^{n-k} ds. \tag{2.45}$$

By Fubini's theorem,

$$\begin{aligned}\int_0^1 \frac{(1-s)^{k-1}\mathbb{E}[N^{(k)}s^{N-k}]}{(k-1)!\mathbb{P}(N\geq k)}ds &= \mathbb{E}\left[\int_0^1 \frac{(1-s)^{k-1}N^{(k)}s^{N-k}}{(k-1)!\mathbb{P}(N\geq k)}ds\right]\\ &= \mathbb{E}\left[\frac{N^{(k)}}{(k-1)!}\int_0^1 (1-s)^{k-1}s^{N-k}ds\middle| N\geq k\right]\\ &= \mathbb{E}\big[1\big|N\geq k\big]=1.\end{aligned}$$

□

Lemma 2.4.3.

$$\mathbb{P}[f(\bar{U})|N_T\geq k] = \frac{\mathbb{P}[N_T^{(k)}]}{(k-1)!\mathbb{P}(N_T\geq k)}\int_0^1 (1-s)^{k-1}\mathbb{Q}^{k,T}\Big[s^{N_T-k}f(\bar{\xi})\Big]ds \tag{2.46}$$

Proof. Recall that $\mathbb{Q}^{k,T}(N_T\geq k)=1$. Hence using Lemma 2.4.1 in the first equality below and the identity (2.45) in the second,

$$\begin{aligned}\mathbb{P}[f(\bar{U})|N_T\geq k] &= \frac{\mathbb{P}[N_T^{(k)}]}{\mathbb{P}(N_T\geq k)}\mathbb{Q}^{k,T}\Big[\frac{1}{N_T^{(k)}}f(\bar{\xi})\Big]\\ &= \frac{\mathbb{P}[N_T^{(k)}]}{\mathbb{P}(N_T\geq k)}\mathbb{Q}^{k,T}\Big[\frac{1}{(k-1)!}\int_0^1 (1-s)^k s^{N_T-k}f(\bar{\xi})ds\Big]\\ &= \frac{\mathbb{P}[N_T^{(k)}]}{\mathbb{P}(N_T\geq k)}\frac{1}{(k-1)!}\int_0^1 (1-s)^k\mathbb{Q}^{k,T}\Big[s^{N_T-k}f(\bar{\xi})\Big]ds\end{aligned}$$

□

2.4.2 Proof of the formula for the finite dimensional distributions of $(\pi_t^{k,L,T})_{t\in[0,T]}$

The partition process $(\pi_t^{k,L,T})_{t\in[0,T]}$ is only defined on the event $\{N_T\geq k\}$. In the interest of making formula lighter, we return to our previous convention in making a slight abuse of notation and writing

$$\mathbb{P}(\pi_t^{k,L,T}\in\cdot)$$

rather than

$$\mathbb{P}(\pi_t^{k,L,T}\in\cdot|N_T\geq k)$$

for events relating to the process $(\pi_t^{k,L,T})_{t\in[0,T]}$. We wrap together our work to prove our formula for the finite dimensional distributions of (π_t^T), under the assumption $\mathbb{P}[L^k]<\infty$. Namely for any partition chain γ_i, and times t_i we now weave things together to show that

$$\mathbb{P}(\pi_{t_1}=\gamma_1,...,\pi_{t_n}=\gamma_n) = \int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N_T\geq k)}\prod_{i=0}^{i=n}\prod_{\Gamma_{i,j}\in\gamma_i}F_{\Delta t_i}^{b_{i,j}}\Big(F_{T-t_{i+1}}(s)\Big)ds. \tag{2.47}$$

Proof of Theorem 2.2.2 under k^{th}-moment assumption. By Theorem 2.3.5, for the functional $f(\xi) = \mathbb{1}\{\pi^k_{t_i} = \gamma_i \ \forall\ i\}$ we have

$$\mathbb{Q}[s^{N_T-k} f(\xi)] = \mathbb{Q}^{k,T}\left[s^{N_T-k}\ \mathbb{1}\left\{\pi^k_{t_i} = \gamma_i \ \forall\ i\right\}\right] = \frac{1}{\mathbb{P}[N_T^{(k)}]} \prod_{i=0}^{i=n} \prod_{\Gamma_{i,j}\in\gamma_i} F^{b_{ij}}_{\Delta t_i}(F_{T-t_{i+1}}(s)) \quad (2.48)$$

Now plug equation (2.48) into equation (2.46). □

Now we prove that we can relax the condition $\mathbb{E}[L^k] < \infty$. Recall that for an offspring generating function $f(s) = \mathbb{E}[s^L]$, the non-explosion hypothesis states that $\int_{1-\epsilon}^1 \frac{ds}{|f(s)-s|} = \infty$ holding for every ϵ near 0 is necessary and sufficient to ensure $\mathbb{P}(N_t < \infty) = 1$ for every t.

In order to safely use a coupling argument in Lemma 2.4.6, we require the following result.

Lemma 2.4.4. Let $\bar{N}_t = |\bigcup_{s\in[0,t]} \mathcal{N}_s|$, the number of individuals who have ever lived by time t. Then, provided the non-explosion hypothesis holds, $\mathbb{P}(\bar{N}_t < \infty) = 1$.

Proof. Suppose we have a continuous-time Galton-Watson process $N := (N_t)_{t\geq 0}$ with offspring generating function $f(s) = \mathbb{E}[s^L] = p_0 + p_1 s + s^2 g(s)$ satisfying the non-explosion hypothesis. Couple N with another process $M := (M_t)_{t\geq 0}$ with generating function $f^*(s) = (p_0+p_1+g(s))s^2$ as follows. Every time an individual in the process N has 0 or 1 children, the corresponding individual in the process M has 2 children. Writing M_t for the number who have ever lived until t in the M-process, clearly $\mathbb{P}(\bar{N}_t \leq \bar{M}_t) = 1$, and it is straightforward to verify that $f^*(s)$ also satisfies the non-explosion hypothesis, and hence M_t is almost surely finite.

Consider in the process M that every individual is replaced by at least two individuals upon death, and hence there can have been at most $\frac{1}{2}M_t$ parents of individuals alive at time t. A similar argument says that there can have been at most $\frac{1}{4}M_t$ grandparents, and so forth. It follows that the we can bound above the number who have ever lived: $\bar{M}_t \leq \sum_{i\geq 0} 2^{-i} M_t = 2M_t$.

Since $2M_t \geq \bar{M}_t \geq \bar{N}_t$, the latter quantity is almost surely finite. □

The following lemma, a variant of the dominated convergence theorem, will be used in the proofs of Lemma 2.4.6, Theorem 2.2.4 and Theorem 2.2.6.

Lemma 2.4.5. Let $g, (g_n)$, and $h, (h_n)$ be measurable functions on $(\Omega, \mathcal{A})$, with $|g_n| \leq h_n$ for all n, and such that $g_n \to g$, $h_n \to h$, and let μ be a finite measure such that $\mu h_n \to \mu h$. Then $\mu g_n \to \mu g$.

Proof. See [25, Theorem 1.21]. □

Lemma 2.4.6. Theorem 2.2.2 holds for every offspring distribution L such that $\mathbb{P}(L \geq 2) > 0$ and satisfying the non-explosion hypothesis.

Proof. Our proof idea as follows. To calculate the distribution of $(\pi^T_t)_{t\in[0,T]}$, first calculate $(\pi^T_t)_{t\in[0,T]}$ from a tree where branch sizes are bounded by n, and hence the moments are finite and the formula (2.14) applies. Then we send $n \to \infty$, showing the formula (2.14) converge suitably.

Let L be a random variable, and let Tree_T be the continuous-time Galton-Watson tree run until time T and $(N_t)_{t\in[0,T]}$ be the corresponding process. We couple the tree Tree_T with a tree with bounded branching as follows. Let $\mathsf{Tree}_{T,n}$ be the tree with offspring distribution $L\mathbb{1}_{L\leq n}$ taken by replacing any birth of size greater than n Tree_T with a birth of size zero, and let $(N_{n,t})_{t\in[0,T]}$ be the associated process.

By Lemma 2.4.4, $\mathbb{P}(\bar{N}_t < \infty) = 1$, and thus $\mathbb{P}(\bar{N}_t \leq n) \uparrow 1$ as $n \to \infty$. Note that $\{\bar{N}_t \leq n\}$ ensures $\{\mathsf{Tree}_{T,n} = \mathsf{Tree}_T\}$, since if at most n particles have ever lived, no particle ever had more than n offspring. It follows that $\mathbb{P}(\mathsf{Tree}_{T,n} = \mathsf{Tree}_T) \uparrow 1$ as $n \to \infty$. In particular, $N_{n,t} \to N_t$ almost surely and hence $\mathbb{P}(N_{n,T} \geq k) \to \mathbb{P}(N_T \geq k)$.

If we pick k individuals from $\mathsf{Tree}_{T,n}$ and call the partition process $(\pi_t^{n,T})_{t\in[0,T]}$, it follows that $\mathbb{P}\Big((\pi_t^{n,T})_{t\in[0,T]} = (\pi_t^T)_{t\in[0,T]}\Big) \to 1$, since the partition processes correspond to subtrees of the trees $\mathsf{Tree}_{n,T}$ and Tree_T respectively.

It remains to check that for a process $(N_{n,t})_{t\geq 0}$ with offspring distribution $L\mathbb{1}_{L\leq n}$ and generating function $F_{n,t}(s)$, that as $n \uparrow \infty$,

$$\int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N_{n,T} \geq k)} \prod_{i=0}^{i=n} \prod_{\Gamma_{i,j}\in\gamma_i} F_{n,\Delta t_i}^{,b_{ij}}\Big(F_{n,T-t_{i+1}}(s)\Big)ds \tag{2.49}$$

$$\to \int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N_T \geq k)} \prod_{i=0}^{i=n} \prod_{\Gamma_{i,j}\in\gamma_i} F_{\Delta t_i}^{b_{ij}}\Big(F_{T-t_{i+1}}(s)\Big)ds. \tag{2.50}$$

First, let us establish that for all (j,t,s), that as $n \to \infty$ $F_{n,t}^j(s) \to F_t^j(s)$.

If $s = 1$, since $N_{n,t} \uparrow N_t$ almost surely, $F_{n,t}^j(1) = \mathbb{E}[N_{n,t}^{(j)}] \uparrow \mathbb{E}[N_t^{(j)}] = F_t^j(1)$ by the monotone convergence theorem.

If $s < 1$, then the function $n \mapsto n^{(j)}s^{n-j}$ is bounded for $n \in \{0,1,2,\ldots\}$ by a constant $M > 0$. Now note that $N_{n,t} \to N_t$ implies $M \geq N_{n,t}s^{N_{n,t}-j} \to N_t s^{N_t - j}$, and hence $F_{n,t}^j(s) = \mathbb{E}[N_{n,t}s^{N_{n,t}-j}] \to \mathbb{E}[N_t s^{N_t-j}] = F_t^j(s)$ by the bounded convergence theorem.

To see the convergence of (2.49) to (2.50), the Faà di Bruno formula for smooth function semigroups, equation (2.20), states that

$$F_T^k(s) = \sum_{(\gamma_1,\ldots,\gamma_n)\in\mathcal{CP}_k^n} \prod_{i=0}^{i=n} \prod_{\Gamma_{i,j}\in\gamma_i} F_{\Delta t_i}^{b_{ij}}\Big(F_{T-t_{i+1}}(s)\Big). \tag{2.51}$$

Since $F_t^j(s) \geq 0$ for every (j,t,s), every term in the sum above is non-negative, and we have the domination relation

$$H(s) := \frac{(1-s)^{k-1}F_T^k(s)}{(k-1)!\mathbb{P}(N_T \geq k)} \geq \frac{(1-s)^{k-1}\prod_{i=0}^{i=n}\prod_{\Gamma_{i,j}\in\gamma_i} F_{\Delta t_i}^{b_{ij}}\Big(F_{T-t_{i+1}}(s)\Big)}{(k-1)!\mathbb{P}(N_T \geq k)} =: G(s) \geq 0. \tag{2.52}$$

Similarly,

$$H_n(s) := \frac{(1-s)^{k-1}F_{n,T}^k(s)}{(k-1)!\mathbb{P}(N_{n,T} \geq k)} \geq \frac{(1-s)^{k-1}\prod_{i=0}^{i=n}\prod_{\Gamma_{i,j}\in\gamma_i} F_{n,\Delta t_i}^{b_{ij}}\Big(F_{n,T-t_{i+1}}(s)\Big)}{(k-1)!\mathbb{P}(N_{n,T} \geq k)} =: G_n(s) \geq 0. \tag{2.53}$$

By our assumption $\mathbb{P}(N_T \geq k) > 0$, and hence by Lemma 2.4.2, for every n, $\int_0^1 H_n(s)ds = 1 = \int_0^1 H(s)ds$. Trivially, $\int_0^1 H_n(s)ds \to \int_0^1 H(s)ds$. So $G_n(s) \to G(s)$ pointwise, $H_n(s) \to H(s)$ pointwise, $\int H_n \to \int H$, and $H_n \geq |G_n|$. It follows by lemma 2.4.5 that $\int G_n \to \int G$. $\square$

2.4.3 Proof of the mixture representation

Here we prove our remark at the end of Section 2.2.1 that $(\pi_t^{k,L,T})_{t\in[0,T]}$ is a mixture of stochastic processes constructed as follows. Sample a random variable S from the probability density

$$\mathbb{P}(S \in ds) = M^{k,L,T}(ds) = \frac{(1-s)^{k-1}F_T^k(s)}{(k-1)!\mathbb{P}(N_T \geq k)}ds, \ s \in [0,1], \tag{2.54}$$

conditional on $S = s$, $(\pi_t^{k,L,T})_{t\in[0,T]}$ has finite dimensional distributions given by

$$K^{k,L,T}(\pi_{t_i} = \gamma_i \ \forall i|s) = \frac{\prod_{i=0}^n \prod_{\Gamma_{i,j}\in\gamma_i} F_{\Delta t_i}^{b_{i,j}} \circ F_{T-t_{i+1}}(s)}{F_T^k(s)} \tag{2.55}$$

Proof. Clearly, with the definitions (2.54) and (2.55) above, we may rewrite (2.14):

$$\begin{aligned}\mathbb{P}(\pi_{t_i}^T = \gamma_i, \ \forall i) &= \int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N_T \geq k)} \prod_{i=0}^{i=n} \prod_{\Gamma_{i,j}\in\gamma_i} F_{\Delta t_i}^{b_{ij}}\Big(F_{T-t_{i+1}}(s)\Big)ds \\ &= \int_0^1 M^{k,L,T}(ds)K^{k,L,T}(\pi_{t_i} = \gamma_i \ \forall i|s).\end{aligned}$$

It remains to see that by Lemma 2.4.2, $M^{k,L,T}$ is a probability measure on $[0,1]$, and that K is a probability measure on the product space $\mathcal{P}_k^{[0,T]}$ by (2.20). $\square$

Remark. In the proof of the split time representation, we will use the mixture representation to sidestep technical issues of integral convergence.

2.4.4 Proof of the split time representation

Recall we say $[\{1,...,k\}] = \eta_0 \prec \eta_1 \prec ... \prec \eta_n = [\{1\},...,\{k\}]$ is a maximal chain of partitions if η_{i+1} is created from η_i by breaking precisely one block of η_i into $c_i \geq 2$ blocks in η_{i+1}, and we write

$$\{\pi : \eta_0 \prec^{dt_1} \eta_1 \prec^{dt_2} ... \prec^{dt_n} \eta_n\} = \bigcap_{i=1}^n \{\pi_{t_i} = \eta_{i-1}, \pi_{t_i+dt_i} = \eta_i\}. \tag{2.56}$$

Let $(\eta_i)_{i=0,..n}$ be a maximal chain where a block of size c_i breaks at each split. Now we prove the split time representation, which states that,

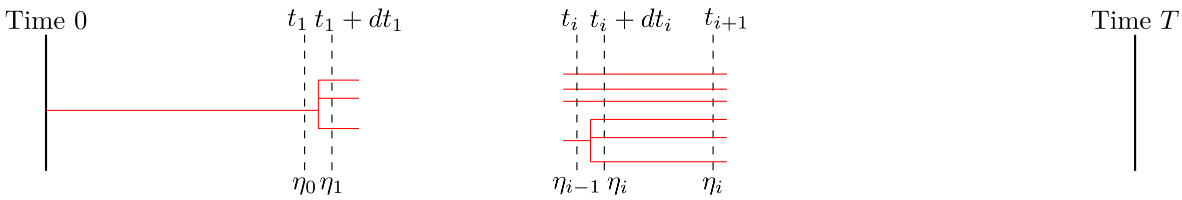
$$\mathbb{P}(\pi^{k,L,T} : \eta_0 \to^{dt_1} \eta_1 \to^{dt_2} ... \to^{dt_n} \eta_n) = \int_0^1 \frac{(1-s)^{k-1}F_T'(s)}{(k-1)!\mathbb{P}(N_T \geq k)} \prod_{i=1}^n \Big(F_{T-t_i}^1(s)^{c_i-1} f^{c_i}(F_{T-t_i}(s))dt_i\Big)ds. \tag{2.57}$$

Proof of theorem 2.2.3. By the mixture representation, (π_t^T) is a mixture of stochastic processes with finite dimensional distributions given by

$$K(\pi_{u_i} = \gamma_i | s) = \frac{\prod_{i=0}^{n} \prod_{\Gamma_{i,j} \in \gamma_i} F_{\Delta u_i}^{b_{i,j}} \circ F_{T-u_{i+1}}(s)}{F_T^k(s)}. \tag{2.58}$$

We calculate the densities for the split times under the measure $K(\cdot|s)$ for $s \in [0,1)$. Namely, for $0 = t_0 < ... < t_n < t_{n+1} = T$, we use (2.14) to give probabilities that the process $(\pi_t^T)_{t\in[0,T]}|S = s$ first visits the state η_i in the time period $(t_i, t_i + dt_i] =: dt_i$. That is, we calculate

$$K(\pi_{t_i} = \eta_{i-1}, \pi_{t_i+dt_i} = \eta_i, \ \forall i | s) \tag{2.59}$$



In order to apply (2.58), let $u_0 = 0, u_{2i-1} = t_i, u_{2i} = t_i + dt_i, i = 1, ..., n, u_{2n+1} = T$, and let $\gamma_{2i} = \gamma_{2i+1} = \eta_i, i = 1, ..., n$.

Let us consider the breakage numbers of the chain $\gamma_0 \prec \gamma_1 \prec ... \prec \gamma_{2n+1}$. During the time period $(t_i, t_i + dt_i]$ we have $\gamma_{2i-1} = \eta_{i-1} \prec \eta_i = \gamma_{2i}$, which means a block of size c_i breaks, and the other $|\eta_{i-1}| - 1$ blocks remain fixed. Hence, since $b_{i-1} = |\eta_{i-1}|$, the associated breakage numbers (up to reordering) of this period are

$$(c_i, \underbrace{1, 1, ..., 1}_{b_{i-1}-1 \text{ times}}).$$

During the period $(t_i + dt_i, t_{i+1}]$, $\gamma_{2i} = \eta_i \prec \eta_i = \gamma_{2i+1}$, so the b_i blocks of η_i remain fixed, hence the breakage numbers associated with this period are

$$(\underbrace{1, 1, ..., 1}_{b_i \text{ times}}).$$

It follows that the numerator of (2.58) is given by

$$F_t^1(F_{T-t}(s)) \prod_{i=1}^{n} \left(F_{dt_i}^{c_i}[F_{T-(t_i+dt_i)}(s)] \ F_{dt_i}^1[F_{T-(t_i+dt_i)}(s)]^{b_{i-1}-1} \right) \ F_{t_{i+1}-(t_i+dt_i)}^1[F_{T-t_{i+1}}(s)]^{b_i} \tag{2.60}$$

Now, it is a straightforward calculation that the generating function $F_t(s) = \mathbb{E}[s^{N_t}]$ satisfies

$$F_h^1(s) = 1 + o(1), \ F_h^n(s) = f^n(s)h + o(h) \text{ for } n \geq 2, \ F_{t-h}^n(s) = F_t^n(s)(1 + o(1)).$$

where $f^n(s) = \mathbb{E}[L^{(n)} s^{L-n}]$. Modulo lower order terms, we can simplify (2.60):

$$F_t'(F_{T-t}(s)) \prod_{i=1}^{n} \left(f^{c_i}[F_{T-t_i}(s)] dt_i \right) \ F_{t_{i+1}-t_i}'[F_{T-t_{i+1}}(s)]^{b_i} \ + \ o(dt_i).$$

Finally, note that $b_i = |\eta_i| = 1 + \sum_{j=1}^{i}(c_i - 1)$. Letting $c_0 := 2$ for convenience, we can collect the F' terms and write

$$F'_t(F_{T-t}(s)) \prod_{i=1}^{n} F'_{t_{i+1}-t_i}[F_{T-t_{i+1}}(s)]^{b_i} = \prod_{i=0}^{n} \Big\{ \prod_{j \geq i} F'_{\Delta t_j}[F_{T-t_{j+1}}(s)] \Big\}^{c_i - 1}.$$

Now by the functional identity

$$F_{T-t_i}(s) = F_{\Delta t_n} \circ ... \circ F_{\Delta t_{i+1}} \circ F_{\Delta t_i},$$

and the chain rule, we have

$$F'_{T-t_i}(s) = \prod_{j \geq i} F'_{\Delta t_j}[F_{T-t_{j+1}}(s)].$$

It then follows that

$$K(\pi_{t_i} = \eta_{i-1}, \pi_{t_i+dt_i} = \eta_i, \ \forall i \ |s) = \frac{\prod_{i=1}^{n} \Big(F^1_{T-t_i}(s)^{c_i - 1} f^{c_i}(F_{T-t_i}(s)) dt_i \Big)}{F^k_T(s)}.$$

□

2.5. Proofs of the supercritical and subcritical asymptotics

2.5.1 Proof of the supercritical asymptotics

Suppose $\mathbb{E}[L] > 1$ and $\mathbb{E}[L \log_+ L] < \infty$. Now we show that $T \to \infty$, $(\pi^{k,L,T}_t)_{t \in [0,T]}$ tends in distribution to the $\mathcal{P}_k$-valued stochastic process $(\bar{\pi}^{k,L}_t)_{t \in [0,\infty)}$, and that $(\bar{\pi}^{k,L}_t)_{t \in [0,\infty)}$ has splitting times given by

$$\begin{aligned}
&\mathbb{P}(\bar{\pi}^{k,L} : \eta_0 \to^{dt_1} \eta_1 \to^{dt_2} ... \to^{dt_n} \eta_n) \\
&= \int_0^\infty \frac{v^{k-1}\varphi^1(v)}{(k-1)!(1-\varphi^0(\infty))} \prod_{i=1}^{n} \Big\{ \Big(\varphi^1(ve^{-(m-1)t_i}) e^{-(m-1)t_i} \Big)^{c_i - 1} f^{c_i}(\varphi^0(ve^{-(m-1)t_i})) dt_i \Big\} dv
\end{aligned} \tag{2.61}$$

where $\varphi^j(v) = \mathbb{E}[W^j e^{-vW}]$ is the generating function of the limit martingale W and $(1 - \varphi^0(\infty))$ is the survival probability.

Proof of Theorem 2.2.4. By the split time representation (2.62), for any $t_1 < ... < t_n \in [0, \infty)$

$$\mathbb{P}(\pi^{k,L,T} : \eta_0 \to^{dt_1} \eta_1 \to^{dt_2} ... \to^{dt_n} \eta_n) = \int_0^1 \frac{(1-s)^{k-1} F'_T(s) ds}{(k-1)!\mathbb{P}(N_T \geq k)} \prod_{i=1}^{n} \Big(F^1_{T-t_i}(s)^{c_i-1} f^{c_i}(F_{T-t_i}(s)) dt_i \Big). \tag{2.62}$$

It follows by Fubini's theorem that for any $t_1 < t_1 + h_1 < t_2 < t_2 + h_2 < ... < t_n + h_n$,

$$\mathbb{P}(\pi^{k,L,T} \in F) := \mathbb{P}(\pi_{t_i} = \eta_{i-1}, \pi_{t_i+h_i} = \eta_i \ \forall \ i = 1, ..., n) \tag{2.63}$$

$$= \int_0^1 \frac{(1-s)^{k-1} F^1_T(s) ds}{(k-1)!\mathbb{P}(N_T \geq k)} \prod_{i=1}^{n} \Big\{ \int_{t_i}^{t_i+h_i} \Big(F^1_{T-u_i}(s)^{c_i-1} f^{c_i}(F_{T-u_i}(s)) \Big) du_i \Big\}. \tag{2.64}$$

We now show that as $T \to \infty$, equation (2.64) converges to

$$\mathbb{P}(\bar{\pi}^{k,L} \in F) := \int_0^\infty \frac{v^{k-1}\varphi^1(v)}{(k-1)!(1-\varphi(\infty))} \Big\{ \int_{t_i}^{t_i+h_i} \Big(e^{-(m-1)u_i}\varphi^1(ve^{-(m-1)u_i})\Big)^{c_i-1} f^{c_i}[\varphi(ve^{-(m-1)u_i})]du_i \Big\} dv$$

from which the splitting times (2.61) can be obtained by differentiating with respect to h_i and then setting $h_i = 0$.

Note that for each v, and for $j \in \{0, 1, 2, ...\}$

$$G^j_{T,u}(v) : = e^{-j(m-1)T} F^j_{T-u}(e^{-ve^{-(m-1)T}})] \tag{2.65}$$

$$= \mathbb{E}\left[\frac{N^{(j)}_{T-u}}{e^{j(m-1)T}} e^{-ve^{-(m-1)T}(N_{T-u}-j)} \right] \tag{2.66}$$

$$\to e^{-j(m-1)u}\mathbb{E}[W^j e^{-ve^{-(m-1)u}W}] \tag{2.67}$$

Equation (2.67) suggests that the change of variable $s \mapsto e^{-ve^{-(m-1)T}}$ captures where the mass lies in the s-integrals of (2.62) and (2.64) for large T. Noting that $\sum_{i=1}^n (c_i - 1) = k - 1$, and multiplying and dividing by suitable powers of $e^{(m-1)T}$ to normalise, we obtain from the change of variable $s \mapsto e^{-ve^{-(m-1)T}}$

$$\mathbb{P}(\pi^{k,L,T} \in F) = \int_0^\infty A_T(v)dv \prod_{i=1}^n \left\{ \int_{t_i}^{t_i+h_i} G^1_{T,u_i}(v)^{c_i-1} f^{c_i}[G^0_{T,u_i}(v)]\ du_i \right\}.$$

where

$$A_T(v) = e^{(k-1)(m-1)T} \frac{(1 - e^{-ve^{-(m-1)T}})^{k-1} G^1_{T,0}(v) e^{-ve^{-(m-1)T}} dv}{(k-1)!\mathbb{P}(N_T \geq k)}.$$

Now by (2.67) and the fact that $(1 - e^{-x}) \sim x$ for small x, for each v

$$A_T(v) \to A(v) := \frac{v^{k-1}\varphi^1(v)}{(k-1)!(1-\varphi(\infty))}.$$

Now we show that for $v > 0$,

$$\int_t^{t+h} G^1_{T,u}(v)^{c-1} f^c[G^0_{T,u}(v)]du \to \int_t^{t+h} \Big(e^{-(m-1)u}\varphi^1(ve^{-(m-1)u})\Big)^{c-1} f^c[\varphi(ve^{-(m-1)u})]du. \tag{2.68}$$

First of all, note by the continuity of f^c on $[0, 1)$, that pointwise in u

$$G^1_{T,u}(v)^{c-1} f^c[G^0_{T,u}(v)] \to \Big(e^{-(m-1)u}\varphi^1(ve^{-(m-1)u})\Big)^{c-1} f^c[\varphi(ve^{-(m-1)u})]. \tag{2.69}$$

Now we show that for fixed v, t, h, c and there exists $M > 0$, and $T_0 > 0$ such that for every $T \geq T_0$, for all $u \in [t, t + h]$

$$G^1_{T,u}(v)^{c-1} f^c[G^0_{T,u}(v)] \leq M$$

establishing (2.68) by the bounded convergence theorem. Note that, with $W_t = N_t e^{-(m-1)t}$,

$$G^1_{T,u}(v) = e^{-(m-1)u}\mathbb{E}[W_{T-u}e^{-ve^{-(m-1)T}(N_{T-u}-j)}] \leq \mathbb{E}[W_{T-u}] = 1$$

Now we show that there exists $M > 0$, and $T_0 > 0$ such that for every $T \geq T_0$, for all $u \in [t, t+h$,

$$f^c[G^0_{T,u}(v)] \leq M.$$

Note that f^c is monotone increasing and if $\mathbb{E}[L^c] = \infty$ then $\lim_{x\uparrow 1} f^c(x) = \infty$. Due to this possibility, it is necessary for us to establish that

$$G^0_{T,u}(v)$$

is uniformly bounded away from 1. W is nondegenerate, hence for each $w > 0$, $\varphi(w) = 1-\epsilon < 1$. Hence there exists $T_1(w)$ such that for every $T \geq T_1$,

$$\mathbb{E}[e^{-wW_T}] \leq 1 - \epsilon/2$$

By setting $w = ve^{-(m-1)(t+h)}$, and setting $T_0 = T_1(w) + (t+h)$, for every $T \geq T_0$, for every $u \in [t, t+h]$,

$$G^0_{T,u}(v) = \mathbb{E}[\exp\left\{-ve^{-(m-1)u}W_{T-u}\right\}] = \mathbb{E}[\exp\{-wW_{T-u}\}] < 1 - \epsilon/2.$$

So by (2.69) and the fact that for each v, $G^1_{T,u}(v)^{c-1}f^c[G^0_{T,u}(v)]$ is uniformly bounded for $u \in [t, t+h]$, (2.68) holds.

So far we have shown that for each $v > 0$,

$$g_T(v) := A_T(v)\prod_{i=1}^{n}\left\{\int_{t_i}^{t_i+h_i} G^1_{T,u_i}(v)^{c_i-1}f^{c_i}[G^0_{T,u_i}(v)]\;du_i\right\}$$
$$\to g(v) := A(v)\prod_{i=1}^{n}\Big\{\int_{t_i}^{t_i+h_i}\left(e^{-(m-1)u_i}\varphi^1(ve^{-(m-1)u_i})\right)^{c_i-1}f^{c_i}[\varphi(ve^{-(m-1)u_i})]du_i\Big\}$$

In order to show $\int_0^\infty g_T(v)dv \to \int_0^\infty g(v)dv$, we use a dominated convergence argument identical to the one used in Lemma 2.4.6. Namely, in light of the mixture representation,

$$\frac{(1-s)^{k-1}F^1_T(s)ds}{(k-1)!\mathbb{P}(N_T \geq k)}\prod_{i=1}^{n}\Big\{\int_{t_i}^{t_i+h_i}\Big(F^1_{T-u_i}(s)^{c_i-1}f^{c_i}(F_{T-u_i}(s))\Big)du_i\Big\} \leq \frac{(1-s)^{k-1}F^k_T(s)ds}{(k-1)!\mathbb{P}(N_T \geq k)},$$

where as we saw in Lemma 2.4.2, the right hand side is a probability density on $[0,1]$. Changing variable $s \mapsto e^{-ve^{-(m-1)T}}$, the domination relation still holds, and under this change of variable, the right hand side

$$h_T(v) := \frac{(1-e^{-ve^{-(m-1)T}})^{k-1}F^k_T(e^{-ve^{-(m-1)T}})}{(k-1)!\mathbb{P}(N_T \geq k)}e^{-ve^{-(m-1)T}}e^{-(m-1)T},\; v \in [0,\infty)$$

is still a probability density on $[0,\infty)$. By (2.67),

$$h_T(v) \to h(v) := \frac{v^{k-1}\mathbb{E}[W^k e^{-(m-1)W}]}{(k-1)!(1-\tilde{\varphi}(\infty))}.$$

To see that $h(v)$ is also a probability density, note by Fubini's theorem that

$$\begin{aligned}\int_0^\infty h(v)dv &= \mathbb{E}\left[\int_0^\infty \frac{v^{k-1}Z^k e^{-vZ}}{\Gamma(k)\mathbb{P}(Z>0)}dv\right]\\ &= \mathbb{E}\left[\frac{\mathbb{1}\{Z>0\}}{\Gamma(k)\mathbb{P}(Z>0)}\int_0^\infty (vZ)^{k-1}e^{-vZ}d(vZ)\right]\\ &= \frac{1}{\mathbb{P}(Z>0)}\mathbb{E}\left[\mathbb{1}\{Z>0\}\right] = 1\end{aligned}$$

Now g_T, g, h_T, h clearly satisfy the setup of Lemma 2.4.5, and hence $\int g_T \to \int g$.

□

2.5.2 Proof of the subcritical asymptotics

Suppose $\mathbb{E}[L] < 1$. Now we show that as $T \to \infty$, $(\rho_t^{k,L,T})_{t\in[0,T]}$ converges in distribution to $(\bar{\rho}_t^{k,L})_{t\in[0,\infty)}$, with split times given by

$$\mathbb{P}(\bar{\rho}^{k,L} : \eta_0 \succ^{dt_1} \eta_1 \succ^{dt_2} ... \succ^{dt_n} \eta_n) = \int_0^1 \frac{(1-s)^{k-1}B'(s)}{(k-1)!\mathbb{P}(W\geq k)}\prod_{i=1}^n \Big\{F'_{t_i}(s)^{c_i-1}f^{c_i}(F_{t_i}(s))dt_i\Big\}ds$$

Proof of Theorem 2.2.6. Picking up from equation (2.64) in the proof of the supercritical case, and replacing t_i with $T-t_i$ and u_i with $T-u_i$, we have

$$\mathbb{P}(\rho_{t_i-h_i} = \eta_{i-1}, \rho_{t_i} = \eta_i \ \forall\ i=1,...,n) \tag{2.70}$$

$$= \int_0^1 \frac{(1-s)^{k-1}F_T^1(s)ds}{(k-1)!\mathbb{P}(N_T\geq k)}\prod_{i=1}^n\Big\{\int_{t_i}^{t_i+h_i}\Big(F_{u_i}^1(s)^{c_i-1}f^{c_i}(F_{u_i}(s))\Big)du_i\Big\}. \tag{2.71}$$

By the conditional limit law for subcritical Galton-Watson processes

$$\lim_{T\to\infty}\mathbb{E}[s^{N_T}|N_T\geq 1] =: B(s)$$

is a probability generating function. Thus

$$\lim_{T\to\infty}\frac{F_T^1(s)}{\mathbb{P}(N_T\geq k)} = \lim_{T\to\infty}\frac{\mathbb{E}[N_T s^{N_t-1}]}{\mathbb{P}(N_T\geq k)} = \lim_{T\to\infty}\frac{\mathbb{P}(N_T\geq 1)}{\mathbb{P}(N_T\geq k)}\mathbb{E}[N_T s^{N_T-1}|N_T\geq 1] = \frac{B'(s)}{\mathbb{P}(W\geq k)}$$

To see that the integrals converge, note that by the mixture representation, we have the domination relation

$$\frac{(1-s)^{k-1}F_T^1(s)ds}{(k-1)!\mathbb{P}(N_T\geq k)}\prod_{i=1}^n\Big\{\int_{t_i}^{t_i+h_i}\Big(F_{u_i}^1(s)^{c_i-1}f^{c_i}(F_{u_i}(s))\Big)du_i\Big\} \leq \frac{(1-s)^{k-1}F_T^k(s)}{(k-1)!\mathbb{P}(N_T\geq k)},$$

and apply the argument in the final paragraph of the proof of Lemma 2.4.6. □

Chapter 3

Critical trees with finite variance

This chapter consists of the paper 'Coalescence in continuous-time Galton-Watson trees' [20], by S.C.Harris, S.G.G.Johnston and M.I.Roberts, submitted to arXiv (arXiv:1703.00299) in September 2017.

The coalescent structure of continuous-time Galton-Watson trees

Simon C. Harris, Samuel G.G. Johnston and Matthew I. Roberts

Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK.

September 4, 2017

Abstract

Take a continuous-time Galton-Watson tree. If the system survives until a large time T, then choose k particles uniformly from those alive. What does the ancestral tree drawn out by these k particles look like? Some special cases are known but we give a more complete answer.

3.1. Introduction

Let L be a random variable taking values in $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$. Consider a continuous-time Galton-Watson tree beginning with one initial particle and branching at rate r with offspring distribution L. We will give more details of the model shortly.

Fix a large time T, and condition on the event that at least k particles are alive at time T. Choose k particles uniformly at random (without replacement) from those alive at time T. These particles, and their ancestors, draw out a smaller tree. The general question that we attempt to answer is: what does this tree look like? This is a fundamental question about Galton-Watson trees; several authors have given answers via interesting and contrasting methods for various special cases, usually when $k = 2$. We aim to give a more complete answer with a unified approach that can be adapted to other situations.

Before explaining our most general results we highlight some illuminating examples. Let $\mathcal{N}_t$ be the set of particles that are alive at time t, and write $N_t = \#\mathcal{N}_t$ for the number of particles that are alive at time t. Let $m = \mathbb{E}[L]$ and for each $j \geq 0$ let $p_j = \mathbb{P}(L = j)$. We assume throughout the article, without further mention, that $p_1 \neq 1$.

On the event $\{N_T \geq 2\}$, choose a pair of particles $(U_T, V_T) \in \mathcal{N}_T$ uniformly at random (without replacement). Then let $\mathcal{S}(T)$ be the last time at which these uniformly chosen particles shared a common ancestor. If $N_T \leq 1$ then set $\mathcal{S}(T) = 0$.

If $p_0 \in [0, 1)$ and $p_2 = 1 - p_0$, then the model is known as a birth-death process. In this case we are able to calculate explicitly the distribution of $\mathcal{S}(T)$ conditional on $\{N_T \geq 2\}$. In particular,

- in the supercritical case when $p_2 > p_0$, the law of $\mathcal{S}(T)$ conditional on $\{N_T \geq 2\}$ converges as $T \to \infty$ to a non-trivial distribution with tail satisfying

$$\lim_{T\to\infty} \mathbb{P}(\mathcal{S}(T) \geq t \,|\, N_T \geq 2) \sim 2r(m-1)te^{-r(m-1)t} \quad \text{as} \quad t \to \infty;$$

- in the subcritical case $p_0 > p_2$, the law of $T - \mathcal{S}(T)$ conditional on $\{N_T \geq 2\}$ converges as $T \to \infty$ to a non-trivial distribution with tail satisfying

$$\lim_{T\to\infty} \mathbb{P}(T - \mathcal{S}(T) \geq t \,|\, N_T \geq 2) \sim \Big(1 - \frac{2p_2}{3p_0}\Big)e^{r(m-1)t} \quad \text{as} \quad t \to \infty.$$

In the critical case we can work more generally.

- If L has any distribution satisfying $m = \mathbb{E}[L] = 1$ and $\mathbb{E}[L^2] < \infty$, then the law of $\mathcal{S}(T)/T$ conditional on $\{N_T \geq 2\}$ converges as $T \to \infty$ to a non-trivial distribution on $[0, 1]$ satisfying

$$\lim_{T\to\infty} \mathbb{P}\Big(\frac{\mathcal{S}(T)}{T} \geq t \,\Big|\, N_T \geq 2\Big) = \frac{2(1-t)}{t^2}\Big(\log\Big(\frac{1}{1-t}\Big) - t\Big).$$

This last result (the critical case) is known: Durrett [13] gave a power series expansion, and Athreya [3] gave a representation in terms of a geometric number of exponential random variables, both of which we will show agree with our explicit formula. Lambert [31] gave a similar formula for a certain critical continuous state branching process. Athreya also mentioned that his expression could alternatively be obtained by using the excursion representation of continuum random trees. This method was also used by Popovic [43], Aldous and Popovic [2], Lambert [32] and Lambert and Popovic [33] to investigate related questions. We give more details of this link in Section 3.3.3.

After this article was released, Lambert constructed a remarkable method for obtaining our formulas from contour processes, which gives another technique for obtaining the results above. We discuss this alternative approach in Section 3.3.3.

Beyond the critical case, we can find a distributional scaling limit when L is “near-critical”. We let the distribution of L depend on T, and write $\mathbb{P}_T$ to signify that the Galton-Watson process now depends on T as a result.

- Suppose that L satisfies $\mathbb{E}_T[L] = 1 + \mu/T + o(1/T)$, $\mathbb{E}_T[L(L-1)] = \beta + o(1)$, and that L^2 is uniformly integrable under $\mathbb{P}_T$. Then the law of $\mathcal{S}(T)/T$ conditional on $\{N_T \geq 2\}$ converges as $T \to \infty$ to a non-trivial distribution on $[0,1]$ satisfying

$$\lim_{T\to\infty} \mathbb{P}_T\Big(\frac{\mathcal{S}(T)}{T} \geq s \,\Big|\, N_T \geq 2\Big) = 2\Big(\frac{e^{r\mu(1-s)}-1}{e^{r\mu(1-s)}-e^{r\mu}}\Big) + 2\frac{(e^{r\mu}-1)(e^{r\mu(1-s)}-1)}{(e^{r\mu(1-s)}-e^{r\mu})^2}\log\Big(\frac{e^{r\mu}-1}{e^{r\mu(1-s)}-1}\Big).$$

O’Connell [39, Theorem 2.3] gave this result by using a diffusion approximation, relating the near-critical process to a time-changed Yule tree, and then adapting the method of Durrett [13] from the critical case.

As mentioned above, these special cases—although they are already interesting in their own right—are just a taster of our general results. The effectiveness and adaptability of our method is demonstrated by the fact that it recovers, in these cases, the results of several separate investigations using different techniques [3, 13, 31, 39]. We now describe our general results very briefly. For any $k \geq 2$, under a weak condition on the moments of L, we sample k particles without replacement at time T and trace back the tree induced by them and their ancestors. It turns out that if we view this tree backwards in time, then the coalescent process thus obtained is topologically the same as Kingman’s coalescent, but has different coalescent rates. We give an explicit joint distribution function for the $k-1$ coalescent times; it turns out that they can be constructed by choosing k independent random variables with a certain distribution and renormalising by the maximum.

In Section 3.2 we give more details of our results, in the same order as above. We follow that with discussion of some of the properties of the scaling limit in Section 3.3. In Section 3.4, we introduce the tools required to prove our results, including a change of measure and a version of Campbell’s formula. We then prove our main result for birth-death processes in Section 3.5, and our main result for near-critical processes in Section 3.6.

3.2. Results

We first describe, in more detail than previously, our basic continuous-time Galton-Watson tree. Under a probability measure $\mathbb{P}$, we begin with one particle, the root, which we give the label $\emptyset$. This particle waits an exponential amount of time $\tau_\emptyset$ with parameter r, and then instantaneously dies and gives birth to some offspring with labels $1, 2, \ldots, L_\emptyset$, where $L_\emptyset$ is an independent copy of the random variable L. To be precise, at the time $\tau_\emptyset$ the particle $\emptyset$ is no longer alive and its offspring are. These offspring then repeat, independently, this behaviour: each particle u waits an independent exponential amount of time with parameter r before dying and giving birth to offspring $u1, u2, \ldots, uL_u$ where L_u is an independent copy of L, and so on. We let $p_j = \mathbb{P}(L = j)$ and $m = \sum_{j=1}^\infty jp_j$. Since we will be using more than one probability measure, we will write $\mathbb{P}[\cdot]$ instead of $\mathbb{E}[\cdot]$ for the expectation operator corresponding to $\mathbb{P}$.

Denote by $\mathcal{N}_T$ the set of all particles alive at time T. For a particle $u \in \mathcal{N}_T$ we let τ_u be the time of its death, and define $\tau_u(T) = \tau_u \wedge T$. If u is an ancestor of v, we write $u \leq v$, and if u is a *strict* ancestor of v (i.e. $u \leq v$ and $u \neq v$) then we write $u < v$. For technical reasons we introduce a graveyard Δ which is not alive (it is not an element of $\mathcal{N}_T$).

For a particle $u \in \mathcal{N}_t$ and $s \le t$, let $u(s)$ be the unique ancestor of u that was alive at time s. For two particles $u, v \in \mathcal{N}_T$, let $\sigma(u,v)$ be the last time at which they shared a common ancestor,

$$\sigma(u,v) = \sup\{t \ge 0 : u(t) = v(t)\}.$$

Now fix $k \in \mathbb{N}$, and at time T, on the event $N_T \ge k$, pick k particles $U_T^1, \ldots, U_T^k$ uniformly at random without replacement from $\mathcal{N}_T$. We let $\mathcal{P}_t^k(T)$ be the partition of $\{1,\ldots,k\}$ induced by letting i and j be in the same block if particles U_T^i and U_T^j shared a common ancestor at time t, i.e. if $\sigma(U_T^i, U_T^j) > t$. We order the elements of $\mathcal{P}_t^k(T)$ by their smallest element.

There are two aspects to the information contained in $\mathcal{P}_t^k(T)$. The first is the topological information; given a collection of blocks, which block will split first, and when it does, what will the new blocks created look like? The second is the times at which the splits occur. We will find that in the models we look at, the topological information is (asymptotically) universal and rather simple to describe, whereas the split times are much more delicate and depend on the parameters of the model. In order to separate out these two aspects, we require some more notation.

Let $\nu_t^k(T)$ be the number of blocks in $\mathcal{P}_t^k(T)$, or equivalently the number of distinct ancestors of $U_T^1 \ldots, U_T^k$ that are alive at time t; that is, $\nu_t^k(T) = \#\{u \in \mathcal{N}_t : u < U_T^i \text{ for some } i \le k\}$.

For $i = 1, \ldots, k-1$ let

$$\mathcal{S}_i^k(T) = \inf\{t \ge 0 : \nu_t^k > i\}.$$

We call $S_1^k(T) \le \ldots \le S_{k-1}^k(T)$ the *split times*. For technical reasons it is often easier to consider the unordered split times; we let $(\tilde{\mathcal{S}}_1^k(T), \ldots, \tilde{\mathcal{S}}_{k-1}^k(T))$ be a uniformly random permutation of $(\mathcal{S}_1^k(T), \ldots, \mathcal{S}_{k-1}^k(T))$.

For $i = 0, \ldots, k-1$ let $P_i^k(T) = \mathcal{P}_{S_i^k}^k(T)$, and let $\mathcal{H} = \sigma(P_0^k(T), \ldots, P_{k-1}^k(T))$, so that $\mathcal{H}$ contains all the topological information about the tree generated by $U_T^1, \ldots U_T^k$, but almost no information about the split times.

3.2.1 Birth-death processes

Fix $\alpha \ge 0$ and $\beta > 0$. Suppose that $r = \alpha + \beta$, $p_0 = \alpha/(\alpha+\beta)$ and $p_2 = \beta/(\alpha+\beta)$, with $p_j = 0$ for $j \ne 0, 2$. This is known as a birth-death process with birth rate β and death rate α. Note that since there are only binary splits, if there are at least k particles alive at time T then when we pick k uniformly at random as above there are always exactly $k-1$ distinct split times. Our first theorem gives an explicit distribution for these split times, in the non-critical case and conditional on $\{N_T \ge k\}$.

Theorem 3.2.1. *Suppose that $\alpha \ne \beta$. For any $s_1, \ldots, s_{k-1} \in (0, T]$, the unordered split times are independent of $\mathcal{H}$ and satisfy*

$$\begin{aligned}
&\mathbb{P}(\tilde{\mathcal{S}}_1^k(T) \ge s_1, \ldots, \tilde{\mathcal{S}}_{k-1}^k(T) \ge s_{k-1} | N_T \ge k) \\
&\quad = \frac{k(E_0 - \alpha/\beta)^k}{(E_0-1)^{k-1}} \left[\frac{1}{(E_0 - \alpha/\beta)} \prod_{i=1}^{k-1} \frac{E_i - 1}{E_i - E_0} + \sum_{j=1}^{k-1} \frac{(E_j - 1)}{(E_j - E_0)^2} \Bigg(\prod_{\substack{i=1 \\ i \ne j}}^{k-1} \frac{E_i - 1}{E_i - E_j} \Bigg) \log\Big(\frac{\beta E_0 - \alpha}{\beta E_j - \alpha} \Big) \right]
\end{aligned}$$

where $E_j = e^{(\beta-\alpha)(T-s_j)}$ for each $j = 1, \ldots, k$ and $s_0 = 0$. Furthermore, the partition process $P_0^k(T), P_1^k(T), \ldots, P_{k-1}^k(T)$ has the following description:

- *if $P_i^k(T)$ contains blocks of sizes $a_1, \ldots, a_{i+1}$, the probability that the next block to split will be block j is $\frac{a_j - 1}{k-i-1}$;*

- if a block of size a splits, it creates two blocks whose (ordered) sizes are l and $a-l$ with probability $1/(a-1)$ for each $l = 1, \dots, a-1$.

The case of the Yule tree, in which $\beta = 1$ and $\alpha = 0$, gives simpler formulas for the split times.

Example 3.2.1 (Yule tree)**.** Suppose that $\alpha = 0$ and $\beta = 1$. Then for any $s \in (0, T]$,

$$\mathbb{P}(\tilde{\mathcal{S}}_1^2(T) \geq s \,|\, N_T \geq 2) = \frac{2(e^{-s} - e^{-T})(e^{-s} - 1 + s)}{(1 - e^{-T})(1 - e^{-s})^2}$$

and for any $s_1, s_2 \in (0, T]$,

$$\begin{aligned}&\mathbb{P}(\tilde{\mathcal{S}}_1^3(T) \geq s_1,\, \tilde{\mathcal{S}}_2^3(T) \geq s_2 \,|\, N_T \geq 3)\\ &= \frac{3(e^{-s_1} - e^{-T})(e^{-s_2} - e^{-T})\big(s_1(1 - e^{-s_2})^2 - s_2(1 - e^{-s_1})^2 + (1 - e^{-s_1})(1 - e^{-s_2})(e^{-s_2} - e^{-s_1})\big)}{(1 - e^{-T})^2(1 - e^{-s_1})^2(1 - e^{-s_2})^2(e^{-s_2} - e^{-s_1})}.\end{aligned}$$

Returning to general $\alpha \neq \beta$, the case $k = 2$, mentioned in the introduction, is of particular interest. Note that when $k = 2$, there is only one split time, so the choice of ordered or unordered is irrelevant. To be consistent with the description in the introduction we write $\mathcal{S}(T) = \mathcal{S}_1^2(T)$. Taking a limit as $T \to \infty$ simplifies the formula significantly, although we have to consider the supercritical and subcritical cases separately.

Example 3.2.2 (Supercritical birth-death, $T \to \infty$)**.** Suppose that $\beta > \alpha$. Then for any $s > 0$,

$$\lim_{T \to \infty} \mathbb{P}(\mathcal{S}(T) \geq s \,|\, N_T \geq 2) = \frac{2e^{-(\beta-\alpha)s}}{(1 - e^{-(\beta-\alpha)s})^2}\big((\beta - \alpha)s - 1 + e^{-(\beta-\alpha)s}\big).$$

Example 3.2.3 (Subcritical birth-death, $T \to \infty$)**.** Suppose that $\alpha > \beta$. Then for any $s > 0$,

$$\lim_{T \to \infty} \mathbb{P}(\mathcal{S}(T) \geq T - s \,|\, N_T \geq 2) = \frac{2\alpha^2}{\beta^2}(e^{(\alpha-\beta)s} - 1)\Big(e^{(\alpha-\beta)s} \log\Big(1 + \frac{\beta}{\alpha e^{(\alpha-\beta)s} - \beta}\Big) - \frac{\beta}{\alpha}\Big).$$

To our knowledge all of these results are new. We note (as Durrett also mentioned in [13]) that in the supercritical case, the time $\mathcal{S}(T)$ is likely to be near 0, whereas in the subcritical case, $\mathcal{S}(T)$ is likely to be near T. This much is to be expected, but the detailed behaviour is perhaps more surprising: as mentioned in the introduction, some elementary calculations using the formulas above show that in the supercritical case,

$$\lim_{T \to \infty} \mathbb{P}(\mathcal{S}(T) \geq s \,|\, N_T \geq 2) \sim 2(\beta - \alpha)s e^{-(\beta-\alpha)s} \quad \text{as} \quad s \to \infty,$$

whereas in the subcritical case,

$$\lim_{T \to \infty} \mathbb{P}(T - \mathcal{S}(T) \geq s \,|\, N_T \geq 2) \sim \Big(1 - \frac{2\beta}{3\alpha}\Big)e^{-(\alpha-\beta)s} \quad \text{as} \quad s \to \infty.$$

We can also give analogous results in the critical case $\alpha = \beta$.

Theorem 3.2.2. Suppose that $\alpha = \beta$. For any $s_1, \dots, s_{k-1} \in (0, T]$ with $s_i \neq s_j$ for $i \neq j$, the unordered split times are independent of $\mathcal{H}$ and satisfy

$$\begin{aligned}&\mathbb{P}(\tilde{\mathcal{S}}_1^k(T)/T \geq s_1, \dots, \tilde{\mathcal{S}}_{k-1}^k(T)/T \geq s_{k-1} \,|\, N_T \geq k)\\ &= k\Big(1 + \frac{1}{\beta T}\Big)^k \Bigg[\frac{1}{1 + 1/T}\prod_{i=1}^{k-1}\Big(1 - \frac{1}{s_i}\Big) + \sum_{j=1}^{k-1} \frac{1 - s_j}{s_j^2}\Bigg(\prod_{\substack{i=1\\ i\neq j}}^{k-1} \frac{1 - s_i}{s_j - s_i}\Bigg) \log\Big(\frac{1 + 1/T}{1 - s_j + 1/T}\Big)\Bigg].\end{aligned}$$

Furthermore, the partition process $P_0^k(T), P_1^k(T), \dots, P_{k-1}^k(T)$ has the following description:

- if $P_i^k(T)$ contains blocks of sizes $a_1, \dots, a_{i+1}$, the probability that the next block to split will be block j is $\frac{a_j-1}{k-i-1}$;
- if a block of size a splits, it creates two blocks whose (ordered) sizes are l and $a-l$ with probability $1/(a-1)$ for each $l = 1, \dots, a-1$.

Example 3.2.4. Suppose that $\alpha = \beta$. Then for any $s > 0$

$$\mathbb{P}(\tilde{\mathcal{S}}_1^2(T)/T \geq s \,|\, N_T \geq 2) = 2\Big(1 + \frac{1}{\beta T}\Big)^2 \Big(\frac{1-s}{s^2}\Big)\Big(\log\Big(\frac{1+1/T}{1-s+1/T}\Big) - \frac{s}{1+1/T}\Big)$$

and for any $s_1, s_2 > 0$,

$$\mathbb{P}(\tilde{\mathcal{S}}_1^3(T)/T \geq s_1,\, \tilde{\mathcal{S}}_2^3(T)/T \geq s_2 \,|\, N_T \geq 3)$$
$$= \frac{3(1+\frac{1}{\beta T})^3(1-s_1)(1-s_2)}{s_1^2 s_2^2 (s_2 - s_1)}\Big[s_2^2 \log\Big(\frac{1-s_1+1/T}{1+1/T}\Big) - s_1^2 \log\Big(\frac{1-s_2+1/T}{1+1/T}\Big) + \frac{s_1 s_2 (s_2 - s_1)}{1+1/T}\Big].$$

We can easily let $T \to \infty$ in these formulas, but in the critical case—and even in near-critical cases—if we are willing to take a scaling limit as $T \to \infty$ then we can work much more generally.

3.2.2 Near-critical processes: a scaling limit

We no longer restrict to birth-death processes; the birth distribution L may take any (non-negative integer) value. In order to consider a scaling limit, we take Galton-Watson processes that are *near-critical*, in that the mean number of offspring is approximately $1 + \mu/T$ for some $\mu \in \mathbb{R}$. We also insist that the variance converges. Conditional on survival to time T, we sample k particles uniformly without replacement, and ask for the structure of the genealogical tree generated by these k particles. In other branching models when the population is kept constant, it has been shown that the resulting coalescent process converges as $T \to \infty$ to Kingman's coalescent [46]. We see something slightly different.

To state our result precisely, we need some more notation. Fix $\mu \in \mathbb{R}$ and $\sigma > 0$. Suppose that for each $T > 0$, the offspring distribution L satisfies

- $\mathbb{E}_T[L] = 1 + \mu/T + o(1/T)$
- $\mathbb{E}_T[L(L-1)] = \sigma + o(1)$
- L^2 is uniformly integrable under $\mathbb{P}_T$.

Theorem 3.2.3 (Near-critical scaling limit)**.** Suppose that the conditions above hold. Then the split times are asymptotically independent of $\mathcal{H}$, and if $\mu \neq 0$, then for any $s_1, \dots, s_{k-1} \in (0,1)$ with $s_i \neq s_j$ for any $i \neq j$,

$$\lim_{T\to\infty} \mathbb{P}_T(\tilde{\mathcal{S}}_1^k(T)/T \geq s_1, \dots, \tilde{\mathcal{S}}_{k-1}^k(T)/T \geq s_{k-1} \,|\, N_T \geq k)$$
$$= k\prod_{i=1}^{k-1} \frac{E_i}{E_i - E_0} + k\sum_{j=1}^{k-1} \frac{E_0 E_j}{(E_j - E_0)^2}\Big(\prod_{\substack{i=1\\ i\neq j}}^{k-1} \frac{E_i}{E_i - E_j}\Big)\log\frac{E_0}{E_j}$$

where $E_j = e^{r\mu(1-s_j)} - 1$ for each $j = 0, \dots, k-1$ and $s_0 = 0$. If $\mu = 0$, then instead

$$\lim_{T\to\infty} \mathbb{P}(\tilde{\mathcal{S}}_1^k(T)/T \geq s_1, \dots, \tilde{\mathcal{S}}_{k-1}^k(T)/T \geq s_{k-1} \,|\, N_T \geq k)$$
$$= k\prod_{i=1}^{k-1} \frac{s_i - 1}{s_i} - k\sum_{j=1}^{k-1} \frac{1-s_j}{s_j^2}\Big(\prod_{\substack{i=1\\ i\neq j}}^{k-1} \frac{1-s_i}{s_j - s_i}\Big)\log(1-s_j).$$

Furthermore, the partition process $P_0^k(T), P_1^k(T), \dots, P_{k-1}^k(T)$ has the following description:

- if $P_i^k(T)$ contains blocks of sizes $a_1, \dots, a_{i+1}$, the probability that the next block to split will be block j converges as $T \to \infty$ to $\frac{a_j - 1}{k-i-1}$;
- if a block of size a splits, with probability tending to 1 it creates two blocks whose (ordered) sizes are l and $a - l$ with probability converging to $\frac{1}{a-1}$ for each $l = 1, \dots, a-1$.

In Theorems 3.2.1 and 3.2.2 we saw that the split times were independent of $\mathcal{H}$. This cannot be the case in Theorem 3.2.3, since two or more split times may be equal with positive probability, an event which is captured by both the split times and the topological information $\mathcal{H}$. However we do see that the split times are *asymptotically* independent, in that $\mathbb{P}_T(A \cap B) \to \mathbb{P}_T(A)\mathbb{P}_T(B)$ for any $A \in \sigma(\mathcal{S}_1^k(T), \dots, \mathcal{S}_{k-1}^k(T))$ and $B \in \mathcal{H}$, which is the best that we can hope for.

In the case that the process is actually critical we recover the following simple formula for the split times.

Example 3.2.5 (Critical processes)**.** Suppose that $\mathbb{E}[L] = 1$ and $\mathbb{E}[L^2] < \infty$. Then for any $s \in (0, 1)$,

$$\lim_{T\to\infty} \mathbb{P}(\mathcal{S}(T)/T \geq s \,|\, N_T \geq 2) = \frac{2(s-1)}{s^2}\big(\log(1-s) + s\big). \tag{3.1}$$

Example 3.2.6 (Near-critical scaling limit, $k = 2$)**.** Suppose that the conditions of Theorem 3.2.3 hold with $\mu \neq 0$. Then for any $s \in (0, 1)$,

$$\lim_{T\to\infty} \mathbb{P}_T(\mathcal{S}(T)/T \geq s \,|\, N_T \geq 2) = 2\Big(\frac{e^{r\mu(1-s)} - 1}{e^{r\mu(1-s)} - e^{r\mu}}\Big) + 2\frac{(e^{r\mu} - 1)(e^{r\mu(1-s)} - 1)}{(e^{r\mu(1-s)} - e^{r\mu})^2}\log\Big(\frac{e^{r\mu} - 1}{e^{r\mu(1-s)} - 1}\Big).$$

Both of these examples are known, but to our knowledge the general formula is not. We give more details in Section 3.3.2.

3.3. Further discussion of the results

In this section we investigate further the scaling limit observed in Theorem 3.2.3. Our aim is to understand the limit, compare it to known results, and to explore other ways of obtaining similar representations; in order to keep the calculations to a reasonable length, at times we will not worry too much about the technical details. We will return to full rigour in Sections 3.4, 3.5 and 3.6, in order to prove our main results.

We work under the conditions of Section 3.2.2: we fix $\mu \in \mathbb{R}$ and $\sigma > 0$, and suppose that for each $T > 0$ the offspring distribution L satisfies

- $\mathbb{E}_T[L] = 1 + \mu/T + o(1/T)$
- $\mathbb{E}_T[L(L-1)] = \sigma + o(1)$
- L^2 is uniformly integrable under $\mathbb{P}_T$.

Theorem 3.2.3 says that the rescaled unordered split times, conditional on at least k particles being alive at time T, converge jointly in distribution to an explicit limit,

$$\Big(\frac{\tilde{\mathcal{S}}_1^k(T)}{T}, \dots, \frac{\tilde{\mathcal{S}}_{k-1}^k(T)}{T}\Big) \xrightarrow{(d)} (\tilde{\mathcal{S}}_1^k, \dots, \tilde{\mathcal{S}}_{k-1}^k).$$

We aim to shed some more light on this limit. First we note that, although the split times (for fixed T) do not usually have a joint density—with positive probability one split time may equal another—their scaling limit *does* have a density. Indeed, from the proof of Theorem 3.2.3 (or by checking directly) we see that this density satisfies (with $s_0 = 0$)

$$f_k(s_1, \ldots, s_{k-1}) = \begin{cases} k(r\mu)^{k-1}(1 - e^{-r\mu}) \displaystyle\int_0^\infty \theta^{k-1} \prod_{i=0}^{k-1} \frac{e^{r\mu(1-s_i)}}{(1 + \theta(e^{r\mu(1-s_i)} - 1))^2} \,\mathrm{d}\theta & \text{if } \mu > 0 \\ k \displaystyle\int_0^\infty \theta^{k-1} \prod_{i=0}^{k-1} \frac{1}{(1 + \theta(1 - s_i))^2} \,\mathrm{d}\theta & \text{if } \mu = 0 \\ k(-1)^k(r\mu)^{k-1}(1 - e^{-r\mu}) \displaystyle\int_0^\infty \theta^{k-1} \prod_{i=0}^{k-1} \frac{e^{r\mu(1-s_i)}}{(1 - \theta(e^{r\mu(1-s_i)} - 1))^2} \,\mathrm{d}\theta & \text{if } \mu < 0. \end{cases}$$

3.3.1 A consistent construction of the scaling limit

The following proposition gives a method for consistently constructing the times $(\tilde{\mathcal{S}}_1^k, \ldots, \tilde{\mathcal{S}}_{k-1}^k)$ in the critical case $\mu = 0$.

Proposition 3.3.1. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables on $(0, \infty)$ with density $(1 + x)^{-2}$. Let $M_k = \max_{i \leq k} X_i$, and choose I such that $X_I = M_k$. For $i \leq k$ define $T_i = 1 - X_i/M_k$. Then $(T_1, \ldots, T_{I-1}, T_{I+1}, \ldots, T_k)$ is equal in distribution to $(\tilde{\mathcal{S}}_1^k, \ldots, \tilde{\mathcal{S}}_{k-1}^k)$ in the critical case $\mu = 0$.

Proof. Of course $\mathbb{P}(M_k \leq \theta) = \mathbb{P}(X_1 \leq \theta)^k$, so $\mathbb{P}(M_k \in \mathrm{d}\theta) = k\mathbb{P}(X_1 \in \mathrm{d}\theta)\mathbb{P}(X_1 \leq \theta)^{k-1}$. Thus

$$\begin{aligned}
&\mathbb{P}(T_1 \in \mathrm{d}s_1, \ldots, T_{k-1} \in \mathrm{d}s_{k-1}) \\
&= \int_0^\infty \mathbb{P}(M_k \in \mathrm{d}\theta)\mathbb{P}(T_1 \in \mathrm{d}s_1, \ldots, T_{k-1} \in \mathrm{d}s_{k-1} | M_k = \theta) \\
&= \int_0^\infty k\mathbb{P}(X_1 \in \mathrm{d}\theta)\mathbb{P}(X_1 \leq \theta)^{k-1}\mathbb{P}(1 - \tfrac{X_1}{\theta} \in \mathrm{d}s_1, \ldots, 1 - \tfrac{X_{k-1}}{\theta} \in \mathrm{d}s_{k-1} | X_1 \leq \theta, \ldots, X_{k-1} \leq \theta) \\
&= \int_0^\infty \frac{k}{(1+\theta)^2}\mathbb{P}(X_1 \leq \theta)^{k-1} \prod_{i=1}^{k-1} \mathbb{P}\Big(1 - \frac{X_i}{\theta} \in \mathrm{d}s_i \Big| X_i \leq \theta\Big) \,\mathrm{d}\theta \\
&= \int_0^\infty \frac{k}{(1+\theta)^2} \prod_{i=1}^{k-1} \mathbb{P}\Big(1 - \frac{X_i}{\theta} \in \mathrm{d}s_i\Big) \,\mathrm{d}\theta \\
&= \int_0^\infty \frac{k}{(1+\theta)^2} \Big(\prod_{i=1}^{k-1} \frac{\theta}{(1 + \theta(1 - s_i))^2} \,\mathrm{d}s_i\Big) \,\mathrm{d}\theta.
\end{aligned}$$

This is exactly the density that we saw for $(\tilde{\mathcal{S}}_1^k, \ldots, \tilde{\mathcal{S}}_{k-1}^k)$ at the start of Section 3.3. □

This result, in particular, clarifies the consistency of the split times. Of course, if we choose $k+1$ particles uniformly without replacement at time T, and then forget one of them, the result should be consistent with choosing k particles originally. This is not immediately obvious from the distribution function given in Theorem 3.2.3, but it follows easily from the construction in Proposition 3.3.1.

We can do something similar when $\mu \neq 0$.

Proposition 3.3.2. Suppose that $\mu \neq 0$. Let $X_1, X_2, \dots$ be a sequence of independent and identically distributed random variables on $(0, \infty)$ with density $(1+x)^{-2}$. Let $M_k = \max_{i \leq k} X_i$, and choose I such that $X_I = M_k$. For $i \leq k$ define

$$T_i = 1 - \frac{1}{r\mu} \log\Big(1 + (e^{r\mu} - 1)\frac{X_i}{M_k}\Big).$$

Then $(T_1, \dots, T_{I-1}, T_{I+1}, \dots, T_k)$ is equal in distribution to $(\tilde{\mathcal{S}}_1^k, \dots, \tilde{\mathcal{S}}_{k-1}^k)$.

Proof. Rather than doing the calculation directly, this follows from Proposition 3.3.1 by noting that making the substitution

$$t_i = \frac{e^{r\mu} - e^{r\mu(1-s_i)}}{e^{r\mu} - 1}$$

in the density f_k recovers the critical case from the non-critical. □

It is interesting to compare this procedure to the coalescent point processes of Lambert and Stadler [35]. Of particular interest is Section 6 of their paper, on whether Kingman's coalescent can be built in a similar way.

3.3.2 Comparison to known formulas

As mentioned in the introduction, the critical case $\mu = 0$ has been investigated by other authors. Athreya [3] gave an implicit description of the distributional limit of $\mathcal{S}(T)/T$. (In fact he worked with discrete-time Galton-Watson processes, but this makes no difference in the limit, and we will continue to use our continuous-time terminology and notation for ease of comparison.) By considering the sizes of the descendancies at time T of particles alive at an earlier time sT, Athreya showed that

$$\lim_{T\to\infty} \mathbb{P}(\mathcal{S}(T)/T < s \,|\, N_T \geq 2) = 1 - E[\phi(G_s)]$$

where G_s satisfies $P(G_s = j) = (1-s)s^{j-1}$ for $j \geq 1$, and

$$\phi(j) = E\left[\frac{\sum_{i=1}^j \eta_i^2}{(\sum_{i=1}^j \eta_i)^2}\right]$$

where $\eta_1, \eta_2, \dots$ are independent exponential random variables of parameter 1.

We check that this description of the scaling limit agrees with our own formula (3.1).

Lemma 3.3.3. With ϕ and G_s as described above,

$$E[\phi(G_s)] = \frac{2(s-1)}{s^2}\big(\log(1-s) + s\big).$$

Proof. Suppose first that we are given $\eta_1, \dots, \eta_j$. Let $\gamma_j = \sum_{i=1}^j \eta_i$, and let U_1 and U_2 be independent uniform random variables on $(0, \gamma_j)$. Then for each l, $(\eta_l / \sum_{i=1}^j \eta_i)^2$ is the probability that both U_1 and U_2 fall within the interval (γ_{l-1}, γ_l). Therefore $(\sum_{i=1}^j \eta_i^2)/(\sum_{i=1}^j \eta_i)^2$ is the probability that for some $l \leq j$, both U_1 and U_2 fall within the interval (γ_{l-1}, γ_l).

Suppose now that we are given only the value of γ_j, and let $\tilde{\gamma}_1, \dots, \tilde{\gamma}_{j-1}$ be a uniform permutation of $\gamma_1, \dots, \gamma_{j-1}$. Since $\gamma_1, \gamma_2, \dots$ can be viewed as the arrival times of a Poisson process of parameter 1, we know that given γ_j, the random variables $\tilde{\gamma}_1, \dots, \tilde{\gamma}_{j-1}, U_1, U_2$ are independent uniform random variables on $(0, \gamma_j)$. Therefore the probability that U_1 and U_2 both fall within the interval $(\tilde{\gamma}_{l-1}, \tilde{\gamma}_l)$ for some l is exactly $2/(j+1)$. Since this does not depend on the value of γ_j, we get immediately that $\phi(j) = 2/(j+1)$.

Summing over the possible values of G_s, we get

$$E[\phi(G_s)] = \sum_{j=1}^{\infty} \frac{2}{j+1}(1-s)s^{j-1} = 2\frac{(1-s)}{s^2}\sum_{j=1}^{\infty}\frac{s^{j+1}}{j+1} = 2\frac{(1-s)}{s^2}\int_0^s \frac{u}{1-u}\,\mathrm{d}u$$
$$= 2\frac{(1-s)}{s^2}\Big(\log\Big(\frac{1}{1-s}\Big) - s\Big) = 2\frac{(s-1)}{s^2}\big(\log(1-s)+s\big). \quad \square$$

Durrett [13] also gave a description of the limit $\mathcal{S}(T)/T$ in the critical case, showing that

$$\lim_{T\to\infty} \mathbb{P}(\mathcal{S}(T)/T > s \,|\, N_T \geq 2) = (1-s)\Bigg(1 + 2\sum_{j=1}^{\infty}\frac{s^j}{j+2}\Bigg).$$

It is easy to expand our formula (3.1) as a power series and check that it agrees with the above. Durrett, in fact, went on to give power series expressions for the distributions of $\mathcal{S}_1^3$ and $\mathcal{S}_2^3$. He further stated that it was "theoretically" possible to calculate distributions of split times for $k > 3$, and also mentioned that he could derive a joint distribution for $\mathcal{S}_1^3$ and $\mathcal{S}_2^3$, again in power series form, but that "we would probably not obtain a useful formula". This makes clear the advantage of our method, which gives explicit formulas for the joint distribution for each k without going through an interative procedure.

O'Connell [39] gave exactly the formula in our Example 3.2.6, the near-critical scaling limit in the case $k = 2$. He also provided a very interesting application to a biologically motivated problem: how long ago did the most recent common ancestor of all humans live?

Lambert [31] (in discrete time) and Le [36] (in continuous time) characterized the distribution of the split time in the case $k = 2$ for quite general Galton-Watson processes in terms of an integral formula involving the generating function. They also investigated the case $k \geq 3$, without giving such explicit formulas for the joint distribution of the split times.

3.3.3 Contour processes and the continuum random tree

Athreya [3] mentioned that his result could alternatively be obtained by representing the limiting random trees with Brownian excursions. We give a non-rigorous discussion of this approach.

It is known that a critical Galton-Watson tree conditioned to survive until time T converges, as $T \to \infty$ (in a suitable topology), to a *continuum random tree*. There is a vast literature, beginning with Aldous [1], on continuum random trees as the scaling limit of various discrete structures. For our rough discussion we can think of drawing our tree, conditioned to survive to time T and renormalised by T, and tracing a contour around it starting from the root and proceeding in a depth-first manner from left to right. The height of that contour process converges as $T \to \infty$ to a Brownian excursion $(B_t)_{t\in[0,\nu]}$ conditioned to reach height 1. It is easy to see that two points $u, v \in [0,\nu]$ correspond to the same "vertex" in the limiting tree if they are at the same height and the excursion between u and v is always above B_u; i.e. $B_u = B_v = \inf_{t\in(u,v)} B_t$. The total population of the tree at time sT corresponds to the local time of the Brownian excursion at level s. Choosing two particles at time T, then, means picking two points on the excursion at height 1 according to the local time measure; and the two particles have a common ancestor at time t if the two points chosen are in the same sub-excursion above height t.

In order to calculate the probability of this last event, we (obviously) need to know a little about Brownian excursions. Excursions, indexed by local time, occur according to a Poisson point process with intensity Lebesgue $\times n$ for some excursion measure n. This measure n satisfies

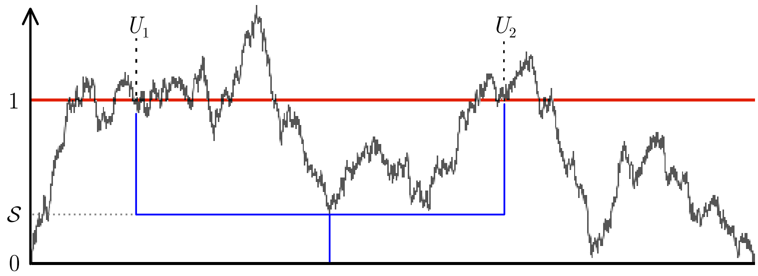


Figure 3-1: *A Brownian excursion conditioned to reach height* 1*. Two points* U_1 *and* U_2 *are chosen uniformly according to local time at height* 1*, and the induced tree is drawn below the excursion. The split time of the two particles is denoted by* $\mathcal{S}$*.*

$n(\sup_t f(t) > a) = \frac{1}{2a}$; and the local time at 0 when the Brownian motion first hits $-\delta$ is exponentially distributed with parameter $\frac{1}{2\delta}$. See for example [44].

Take a Brownian excursion conditioned to reach height 1, and choose two points U_1 and U_2 at height 1 uniformly according to local time measure. Let L_1 be the total local time at level 1, and L_U be the total local time between U_1 and U_2. The event that U_1 and U_2 are in the same sub-excursion above height s is exactly the event that there is no excursion from level 1 between U_1 and U_2 that goes below level s (and stays above level 0); by the facts about Brownian excursions above, given L_U, the number of such excursions is a Poisson random variable with parameter $L_U(\frac{1}{2(1-s)} - \frac{1}{2})$. Thus the probability that U_1 and U_2 are in the same sub-excursion above height s is

$$\int_0^\infty \mathbb{P}(L_1 \in \mathrm{d}x) \int_0^x \mathbb{P}(L_U \in \mathrm{d}y \,|\, L_1 = x) e^{-y(\frac{1}{2(1-s)} - \frac{1}{2})}.$$

The local time L_1 is exponential of parameter $1/2$, and it is easy to check that the density of the distance between two uniform random variables on $(0, x)$ is $2(x-y)/x^2$. Thus the above equals

$$\int_0^\infty \frac{1}{2} e^{-x/2} \int_0^x \frac{2(x-y)}{x^2} e^{-y(\frac{1}{2(1-s)} - \frac{1}{2})} \,\mathrm{d}y\,\mathrm{d}x.$$

Making the substitution $z = y/x$ and changing the order of integration, we get

$$\int_0^1 (1-z) \int_0^\infty e^{-\frac{1}{2}x(1+\frac{z}{1-s}-z)} \,\mathrm{d}x\,\mathrm{d}z,$$

and it is then easy to integrate directly to get that the limiting split time $\mathcal{S}$ satisfies

$$\mathbb{P}(\mathcal{S} \geq s) = 2\Big(\frac{s-1}{s^2}\Big)(\log(1-s) + s)$$

which agrees with (3.1).

Applying this sophisticated machinery works well (at least if we do not worry too much about the technical details) in this simple case. However it becomes much more difficult to generalise these techniques to obtain the joint distribution of the split times for three particles, rather than just two; let alone the general formula for k particles that appeared in Theorem 3.2.3.

Popovic [43] used the following observation. Condition on the event that there are exactly k particles alive at time T_k, so that the k particles we choose comprise the whole population, then rescale by T_k and let $k \to \infty$. If $T_k/k \to t$, then the contour process converges to a Brownian excursion conditioned to have local time 1 at level t; and the split times are then governed by the entire collection of excursions below level t. These excursions form a Poisson point process with an explicit intensity measure. This allowed Popovic to give some very interesting results about critical processes, and similar techniques were built upon in various ways by her and other authors [2, 16, 32, 33]. Although these are certainly related to our investigation, they often look at the entire population alive at time T, rather than sampling a fixed number of individuals, which results in a different scaling regime. Biological motivation for why we might like to sample a fixed number of individuals from a growing population—that is, our regime—can be found in [39].

After this article was released, Lambert [private communication] constructed a remarkable method for obtaining our formulas from contour processes. Given a branching process whose population at time T is geometrically distributed (for example a birth-death process), the work in [35] allows one to sample each particle at time T independently with some fixed probability $y \in (0, 1)$ and reconstruct the genealogical tree of the sampled particles. By taking y to be a realisation of a carefully chosen improper random variable Y, and conditioning the resulting number of particles sampled to be exactly k, Lambert can produce our Proposition 3.5.2. We stress however that finding the correct (improper) distribution for Y would have been extremely difficult without prior knowledge of the answers provided by our results.

Lambert's method works for a large class of processes with geometrically distributed population sizes, known as *coalescent point processes*. For Galton-Watson processes the geometric condition restricts us to birth-death processes, although it is also possible with further work to obtain results in the near-critical regime; see [39].

Another advantage of our approach is that it does not require a Markovian contour process, and has the potential to be generalised for example to Galton-Watson processes with infinite variance, or spatial branching processes. We plan to carry out some of these generalisations in future work.

3.3.4 Purple trees

For a moment forget about the scaling limit, and consider a birth-death process (that is, fix $\alpha \geq 0$ and $\beta > 0$, and suppose that $r = \alpha + \beta$, $p_0 = \alpha/(\alpha + \beta)$ and $p_2 = \beta/(\alpha + \beta)$, with $p_j = 0$ for $j \neq 0, 2$). Wait until time T, and then colour any particle that has a descendant alive at time T purple, and any particle whose descendants all die before time T red.

To put this into context, Harris, Hesse and Kyprianou [22] considered a supercritical branching process and coloured any particle whose descendants survived forever blue, and anyone whose descendants all died out red. We are not interested in whether particles survive forever, only whether they survive to time T, so we colour such particles purple. Of course red particles in our picture are also red in the Harris-Hesse-Kyprianou picture, whereas each of our purple particles may be either red or blue in their colouring.

Now suppose that, rather than running the birth-death process until time T and then colouring all the particles, we want to construct the coloured picture dynamically as the process evolves. If we start with one particle and condition on the process surviving until time T, then the first particle is certainly purple, since at least one of its descendants must survive.

Let $p_t = \mathbb{P}(N_t = 0)$. Using generating functions one can show that

$$p_t = \frac{\alpha e^{(\beta-\alpha)t} - \alpha}{\beta e^{(\beta-\alpha)t} - \alpha}, \qquad 1 - p_t = \frac{(\beta - \alpha)e^{(\beta-\alpha)t}}{\beta e^{(\beta-\alpha)t} - \alpha};$$

see Section 3.5.1 for details.

If a purple particle branches at time s, then its two children could be either both purple, or one red and one purple. The probability that they are both purple must be

$$\frac{(1-p_{T-s})^2}{1-p_{T-s}^2},$$

corresponding to the probability that *both* descendancies survive given that at least one does. The probability that one is purple and one is red must similarly be

$$\frac{2p_{T-s}(1-p_{T-s})}{1-p_{T-s}^2}.$$

One can check from [22] that purple particles branch at rate $\beta(1+p_{T-s})$ at time s, and red particles branch at rate βp_{T-s} at time s. In particular purple particles give birth to new purple particles at rate

$$\beta(1+p_{T-s})\cdot\frac{(1-p_{T-s})^2}{1-p_{T-s}^2}=\beta(1-p_{T-s}).$$

Similar calculations can be done generally, rather than just for birth-death processes. However it is easy to see that in near-critical cases the probability that a purple particle has more than two purple children at any branching event will tend to zero, so in a sense the important information is captured by the simpler birth-death calculations. Indeed we saw in Theorem 3.2.3 that in our scaling limit, only the mean of the branching process really matters; and we will see again in Lemma 3.6.6 that only binary splits appear in the limit. For this non-rigorous discussion we therefore carry out our calculations only in the birth-death case.

Of course, to understand the coalescent structure of the tree drawn out by particles chosen at time T, we can ignore the red particles; only the purple tree matters. Let us now return to a near-critical scaling limit by assuming that $\beta=\alpha+\gamma/T$ for some $\gamma\neq 0$. At time sT, the purple tree branches at rate

$$\beta(1-p_{T-s})=\frac{\beta\gamma e^{\gamma(1-s)}/T}{\beta e^{\gamma(1-s)}-(\beta-\gamma/T)}=\frac{\gamma e^{\gamma(1-s)}}{T(e^{\gamma(1-s)}-1)}\cdot\Big(1-\frac{\gamma}{\beta T(e^{\gamma(1-s)}-1+\frac{\gamma}{\beta T})}\Big).$$

Scaling time onto $[0,1]$, at time $s\in(0,1)$ the purple tree branches at rate

$$\frac{\gamma e^{\gamma(1-s)}}{e^{\gamma(1-s)}-1}.$$

Since

$$\int_0^t\frac{\gamma e^{\gamma(1-s)}}{e^{\gamma(1-s)}-1}\,\mathrm{d}s=\int_{e^{\gamma(1-t)}}^{e^{\gamma}}\frac{1}{u-1}\,\mathrm{d}u=\log\Big(\frac{e^{\gamma}-1}{e^{\gamma(1-t)}-1}\Big),$$

we see that the purple tree in the near-critical scaling limit is the same as a Yule tree (binary branching at rate 1) observed under the time change

$$t\mapsto\log\Big(\frac{e^{\gamma}-1}{e^{\gamma(1-t)}-1}\Big).$$

Following the same route in the purely critical case $\alpha=\beta$ gives that the rescaled purple tree branches at rate $1/(1-s)$, which corresponds to a Yule tree under the time change $t\mapsto-\log(1-t)$.

These rough calculations help to explain the similarities between our formulas in the near-critical scaling limit (Theorem 3.2.3) and in the birth-death process (Theorem 3.2.1).

3.4. Spines and changes of measure

In this section we lay down many of the technical tools that we will need to prove the results in the previous sections. Our two most important signposts will be Proposition 3.4.2, which translates questions about uniformly chosen particles under $\mathbb{P}$ into calculations under a new measure $\mathbb{Q}$; and Proposition 3.4.10, which is a version of Campbell's formula under $\mathbb{Q}$ which will be central to our analysis.

First, of course, we must introduce $\mathbb{Q}$, and we begin by describing the idea of *spines*, which introduce extra information into our tree by allocating *marks* to certain special particles. Spine methods are now well known and a thorough treatment can be found for example in [21]. We give only a brief introduction.

3.4.1 The k-spine measure $\mathbb{P}^k$

We define a new measure $\mathbb{P}^k$ under which there are k distinguished lines of descent, which we call spines. Briefly, $\mathbb{P}^k$ is simply an extension of $\mathbb{P}$ in that all particles behave as in the original branching process; the only difference is that some particles carry marks showing that they are part of a spine.

Under $\mathbb{P}^k$ particles behave as follows:

- We begin with one particle which carries k marks $1, 2, \ldots, k$.
- We think of each of the marks $1, \ldots, k$ as distinguishing a particular line of descent or "spine", and define ξ_t^i to be the label of whichever particle carries mark i at time t.
- A particle carrying j marks $b_1 < b_2 < \ldots < b_j$ at time t branches at rate r, dying and being replaced by a random number of particles according to the law of L, independently of the rest of the system, just as under $\mathbb{P}$.
- Given that a particles $v_1, \ldots, v_a$ are born at a branching event as above, the j marks each choose a particle to follow independently and uniformly at random from amongst the a available. Thus for each $1 \le l \le a$ and $1 \le i \le j$ the probability that v_l carries mark b_i just after the branching event is $1/a$, independently of all other marks.
- If a particle carrying $j > 0$ marks $b_1 < b_2 < \ldots < b_j$ dies and is replaced by 0 particles, then its marks are transferred to the graveyard Δ.

Again we emphasise that under $\mathbb{P}^k$, the system behaves exactly as under $\mathbb{P}$ except that some particles carry extra marks showing the lines of descent of k spines. We write $\xi_t = (\xi_t^1, \ldots, \xi_t^k)$. Obviously ξ_t depends on k too, but we omit this from the notation.

We let n_t be the number of distinct spines (i.e. the number of particles carrying marks) at time t, and for $i \ge 1$

$$\psi_i = \inf\{t \ge 0 : n_t \notin \{1, \ldots, i\}\}$$

with $\psi_0 = 0$. We view ψ_i as the ith spine split time (although, for example, the first and second spine split times may be equal—corresponding to marks following three different particles at the first branching event). We also let ρ_t^i be the number of marks following spine i.

The set of distinct spine particles at any time t, and the marks that are following those spine particles, induce a partition $\mathcal{Z}_t^k$ of $\{1, \ldots, k\}$. That is, i and j are in the same block of $\mathcal{Z}_t^k$ if $\xi_t^i = \xi_t^j$. If we then let

$$Z_i^k = \mathcal{Z}_{\psi_i}^k$$

for $i = 0, \ldots, k-1$, we have created a discrete collection of partitions $Z_0, Z_1, \ldots, Z_{k-1}$ which describe the topological information about the spines without the information about the spine split times. It will occasionally be useful to use the σ-algebra $\mathcal{H}' = \sigma(Z_0, Z_1, \ldots)$.

For any particle $u \in \mathcal{N}_t$, there exists a last time at which u was a spine (which may be t). If this time equals ψ_i for some i, then we say that u is a *residue* particle; if it does not equal ψ_i for any i, and u is not a spine, then we say that u is *ordinary*. Each particle is exactly one of residue, ordinary, or a spine.

Of course $\mathbb{P}^k$ is not defined on the same σ-algebra as $\mathbb{P}$. We let $\mathcal{F}_t^k$ be the filtration containing all information about the system, including the k spines, up to time t; then $\mathbb{P}^k$ is defined on $\mathcal{F}_\infty^k$. For more details see [21, Section 5]. Let $\mathcal{F}_t^0$ be the filtration containing only the information about the Galton-Watson tree. Let $\tilde{\mathcal{G}}_t^k$ be the filtration containing all the information about the k spines (including the birth events along the k spines) up to time t, but none of the information about the rest of the tree. Finally let $\mathcal{G}_t^k$ be the filtration containing information only about spine splitting events (including which marks follow which spines); $\mathcal{G}_t^k$ does not know when births of ordinary particles from the spines occur.

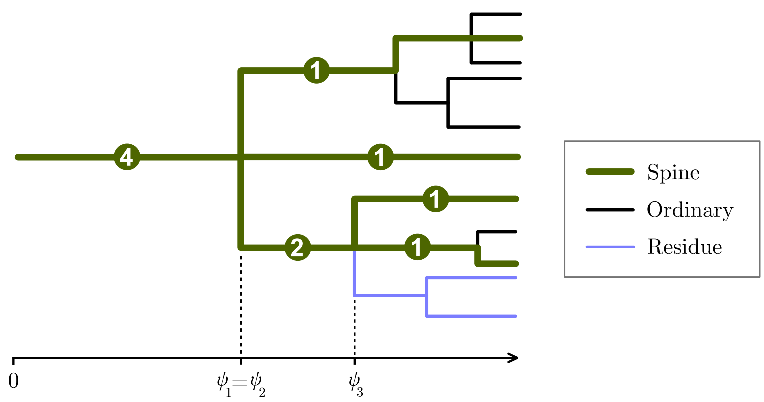


Figure 3-2: *Spines, ordinary particles and residue particles. The horizontal axis represents time. The numbers show how many marks are carried by each spine.*

3.4.2 A change of measure

Throughout the rest of this section we fix $k \geq 1$ and assume that $\mathbb{P}[L^k] < \infty$. This condition will be relaxed later, but it is required even to define our changed measure.

For any set S and $k \geq 1$, let $S^{(k)}$ be the set of distinct k-tuples from S, and for $n \geq 0$, write

$$n^{(k)} = \begin{cases} n(n-1)(n-2)\ldots(n-k+1) & \text{if } n \geq k \\ 0 & \text{otherwise.} \end{cases}$$

Note that $|S^{(k)}| = |S|^{(k)}$. For $t \geq 0$, define

$$g_{k,t} = \mathbb{1}_{\{\xi_t^i \neq \xi_t^j \ \forall i \neq j\}} \prod_{i=1}^{k} \prod_{v < \xi_t^i} L_v$$

and

$$\zeta_{k,t} = \frac{g_{k,t}}{\mathbb{P}[N_t^{(k)}]}.$$

Lemma 3.4.1. For any $t \geq 0$,

$$\mathbb{P}^k[g_{k,t}|\mathcal{F}_t^0] = N_t^{(k)}.$$

In particular, $\mathbb{P}^k[\zeta_{k,t}] = 1$.

Proof.

$$\begin{aligned}\mathbb{P}^k[g_{k,t}|\mathcal{F}_t^0] &= \mathbb{P}^k\Big[\sum_{u\in\mathcal{N}_t^{(k)}} \mathbb{1}_{\{\xi_t=u\}} \prod_{i=1}^k \prod_{v<u_i} L_v \,\Big|\, \mathcal{F}_t^0\Big] \\ &= \sum_{u\in\mathcal{N}_t^{(k)}} \Big(\prod_{i=1}^k \prod_{v<u_i} L_v\Big)\mathbb{P}^k(\xi_t = u \,|\, \mathcal{F}_t^0).\end{aligned}$$

Recall that the marks act independently, and at each branching event choose uniformly amongst the available children. Therefore

$$\mathbb{P}^k(\xi_t = u \,|\, \mathcal{F}_t^0) = \prod_{i=1}^k \mathbb{P}^k(\xi_t^i = u_i \,|\, \mathcal{F}_t^0) = \prod_{i=1}^k \prod_{v<u_i} \frac{1}{L_v}. \tag{3.2}$$

Thus

$$\mathbb{P}^k[g_{k,t}|\mathcal{F}_t^0] = \sum_{u\in\mathcal{N}_t^{(k)}} 1 = |\mathcal{N}_t^{(k)}| = N_t^{(k)}.$$

This gives the first part of the result, and taking expectations gives the second. □

We now fix $T > 0$ and define a new probability measure $\mathbb{Q}^{k,T}$ by setting

$$\left.\frac{\mathrm{d}\mathbb{Q}^{k,T}}{\mathrm{d}\mathbb{P}^k}\right|_{\mathcal{F}_T^k} = \zeta_{k,T}.$$

Often, when the choice of T and k is clear, we write $\mathbb{P}$ instead of $\mathbb{P}^k$ (since $\mathbb{P}^k$ is an extension of $\mathbb{P}$ this should not cause any problems) and $\mathbb{Q}$ instead of $\mathbb{Q}^{k,T}$. We let

$$Z_{k,T} = \frac{N_T^{(k)}}{\mathbb{P}[N_T^{(k)}]}$$

so that, by Lemma 3.4.1,

$$\left.\frac{\mathrm{d}\mathbb{Q}^{k,T}}{\mathrm{d}\mathbb{P}^k}\right|_{\mathcal{F}_T^0} = Z_{k,T}. \tag{3.3}$$

The rest of this section is devoted to understanding the measure $\mathbb{Q}^{k,T}$ and how it might be useful to us.

3.4.3 First properties of $\mathbb{Q}^{k,T}$

Our main aim in this section is to prove the following proposition, which translates questions about particles sampled uniformly without replacement under $\mathbb{P}$ into questions about the spines under $\mathbb{Q}$.

Proposition 3.4.2. Suppose that f is a measurable functional of k-tuples of particles at time T. Then

$$\mathbb{P}\Big[\frac{1}{N_T^{(k)}}\sum_{u\in\mathcal{N}_T^{(k)}} f(u)\,\Big|\, N_T\geq k\Big] = \frac{\mathbb{P}[N_T^{(k)}]}{\mathbb{P}(N_T\geq k)(k-1)!}\int_0^\infty (e^z-1)^{k-1}\mathbb{Q}^{k,T}\Big[e^{-zN_T}f(\xi_T)\Big]\,\mathrm{d}z.$$

Before we prove this, we develop several partial results along the way. The following simple general lemma will be useful.

Lemma 3.4.3. Suppose that μ and ν are probability measures on the σ-algebra $\mathcal{F}$, and that $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$. If

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\bigg|_{\mathcal{F}} = Y \quad\text{ and }\quad \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\bigg|_{\mathcal{G}} = Z,$$

then for any non-negative $\mathcal{F}$-measurable X,

$$Z\mu[X|\mathcal{G}] = \nu[XY|\mathcal{G}] \quad \nu\text{-almost surely.}$$

Proof. For any $A\in\mathcal{G}$,

$$\nu[XY\mathbb{1}_A] = \mu[X\mathbb{1}_A] = \mu[\nu[X|\mathcal{G}]\mathbb{1}_A] = \nu[Z\nu[X|\mathcal{G}]\mathbb{1}_A].$$

Since $Z\mu[X|\mathcal{G}]$ is $\mathcal{G}$-measurable, it therefore satisfies the definition of conditional expectation of XY with respect to $\mathcal{G}$ under ν. $\square$

Applying this to our situation, we get that for any non-negative $\mathcal{F}_T^k$-measurable random variable X, on the event $Z_{k,T}>0$,

$$\mathbb{Q}^{k,T}[X|\mathcal{F}_T^0] = \frac{1}{Z_{k,T}}\mathbb{P}^k[X\zeta_{k,T}|\mathcal{F}_T^0], \tag{3.4}$$

and on the event $\zeta_{k,T}>0$, since $\zeta_{k,T}$ is $\tilde{\mathcal{G}}_T^k$-measurable,

$$\mathbb{Q}^{k,T}[X|\tilde{\mathcal{G}}_T^k] = \frac{1}{\zeta_{k,T}}\mathbb{P}^k[X\zeta_{k,T}|\tilde{\mathcal{G}}_T^k] = \mathbb{P}^k[X|\tilde{\mathcal{G}}_T^k]. \tag{3.5}$$

This last equation (3.5) tells us in particular that any event that is independent of $\tilde{\mathcal{G}}_T^k$ has the same probability under $\mathbb{Q}$ as under $\mathbb{P}$. In other words, non-spine particles behave under $\mathbb{Q}$ exactly as they do under $\mathbb{P}$: they branch at rate r and have offspring distribution L.

Also note that under $\mathbb{Q}^{k,T}$, the k spine particles are almost surely distinct at time T, since directly from the definition of $\zeta_{k,T}$,

$$\mathbb{Q}^{k,T}(\exists i\neq j:\xi_T^i=\xi_T^j) = \mathbb{P}[\zeta_{k,T}\mathbb{1}_{\{\exists i\neq j:\xi_T^i=\xi_T^j\}}] = 0.$$

In fact, the next lemma tells us that under $\mathbb{Q}^{k,T}$, the spines are chosen uniformly without replacement from those alive at time T.

Lemma 3.4.4. For any $u \in \mathcal{N}_T^{(k)}$, on the event $N_T \geq k$,

$$\mathbb{Q}^{k,T}(\xi_T = u | \mathcal{F}_T^0) = \frac{1}{N_T^{(k)}}.$$

Proof. Note that if $N_T \geq k$ then $Z_{k,T} > 0$. Then by (3.4), for any $u \in \mathcal{N}_T^{(k)}$,

$$\mathbb{Q}(\xi_T = u | \mathcal{F}_T^0) = \frac{1}{Z_{k,T}} \mathbb{P}[\zeta_{k,T} \mathbb{1}_{\{\xi_T = u\}} | \mathcal{F}_T^0] = \frac{\mathbb{P}[N_T^{(k)}]}{N_T^{(k)}} \frac{1}{\mathbb{P}[N_T^{(k)}]} \Big(\prod_{i=1}^{k} \prod_{v < u_i} L_v \Big) \mathbb{P}(\xi_t = u | \mathcal{F}_T^0).$$

The result now follows by applying (3.2). □

As part of proving Proposition 3.4.2 we will need to calculate quantities like $\mathbb{Q}[1/N_T^{(k)} | \mathcal{G}_T^k]$. The next lemma allows us to work with moment generating functions, which are somewhat easier to deal with.

Lemma 3.4.5. For any $k \in \mathbb{N}$ and $T \geq 0$,

$$\mathbb{Q}^{k,T}\Big[\frac{1}{N_T^{(k)}} \Big| \mathcal{G}_T^k \Big] = \frac{1}{(k-1)!} \int_0^\infty (e^z - 1)^{k-1} \mathbb{Q}^{k,T}[e^{-zN_T} | \mathcal{G}_T^k] \,\mathrm{d}z.$$

Proof. We show, by induction on j, that for all $j = 1, \ldots, k$,

$$\mathbb{Q}^{k,T}\Big[\frac{1}{N_T^{(j)}} \Big| \mathcal{G}_T^k \Big] = \frac{1}{(j-1)!} \int_0^\infty (e^z - 1)^{j-1} \mathbb{Q}^{k,T}[e^{-zN_T} | \mathcal{G}_T^k] \,\mathrm{d}z.$$

For $j = 1$, by Fubini's theorem,

$$\mathbb{Q}^{k,T}\Big[\frac{1}{N_T} \Big| \mathcal{G}_T^k \Big] = \mathbb{Q}^{k,T}\Big[\int_0^\infty e^{-zN_T} \mathrm{d}z \Big| \mathcal{G}_T^k \Big] = \int_0^\infty \mathbb{Q}^{k,T}[e^{-zN_T} | \mathcal{G}_T^k] \,\mathrm{d}z.$$

For the general step, observe that for $j \leq k - 1$,

$$\begin{aligned}
&\int_0^\infty (e^z - 1)^j \mathbb{Q}^{k,T}[e^{-zN_T} | \mathcal{G}_T^k] \,\mathrm{d}z \\
&\qquad = \int_0^\infty (e^z - 1)^{j-1} \mathbb{Q}^{k,T}[e^{-z(N_T - 1)} | \mathcal{G}_T^k] \mathrm{d}z - \int_0^\infty (e^z - 1)^{j-1} \mathbb{Q}^{k,T}[e^{-zN_T}] \,\mathrm{d}z
\end{aligned}$$

and by the induction hypothesis, this equals

$$\begin{aligned}
&(j-1)! \mathbb{Q}^{k,T}\Big[\frac{1}{(N_T - 1)^{(j)}} \Big| \mathcal{G}_T^k \Big] - (j-1)! \mathbb{Q}^{k,T}\Big[\frac{1}{N_T^{(j)}} \Big| \mathcal{G}_T^k \Big] \\
&= (j-1)! \mathbb{Q}^{k,T}\Big[\frac{N_T}{N_T^{(j+1)}} - \frac{N_T - j}{N_T^{(j+1)}} \Big| \mathcal{G}_T^k \Big] \\
&= j! \mathbb{Q}^{k,T}\Big[\frac{1}{N_T^{(j+1)}} \Big| \mathcal{G}_T^k \Big].
\end{aligned}$$

This gives the result. □

Proof of Proposition 3.4.2. First note that

$$\mathbb{Q}[f(\xi_T)|\mathcal{F}_T^0]\mathbb{1}_{\{N_T\geq k\}} = \mathbb{Q}\Big[\sum_{u\in\mathcal{N}_T^{(k)}} \mathbb{1}_{\{\xi_T=u\}}f(u)\,\Big|\,\mathcal{F}_T^0\Big] = \sum_{u\in\mathcal{N}_T^{(k)}} f(u)\mathbb{Q}(\xi_T=u|\mathcal{F}_T^0)$$

almost surely. Applying Lemma 3.4.4, we get

$$\mathbb{Q}[f(\xi_T)|\mathcal{F}_T^0]\mathbb{1}_{\{N_T\geq k\}} = \frac{1}{N_T^{(k)}}\sum_{u\in\mathcal{N}_T^{(k)}} f(u)$$

almost surely (where we take the right-hand side to be zero if $N_T < k$). Taking $\mathbb{P}$-expectations,

$$\mathbb{P}\Big[\frac{1}{N_T^{(k)}}\sum_{u\in\mathcal{N}_T^{(k)}} f(u)\Big] = \mathbb{P}\big[\mathbb{Q}[f(\xi_T)|\mathcal{F}_T^0]\mathbb{1}_{\{N_T\geq k\}}\big].$$

Applying (3.3) and recalling that under $\mathbb{Q}$ there are at least k particles alive at time T almost surely,

$$\mathbb{P}\Big[\frac{1}{N_T^{(k)}}\sum_{u\in\mathcal{N}_T^{(k)}} f(u)\Big] = \mathbb{Q}\Big[\frac{1}{Z_{k,T}}\mathbb{Q}[f(\xi_T)|\mathcal{F}_T^0]\Big] = \mathbb{Q}\Big[\frac{1}{Z_{k,T}}f(\xi_T)\Big] = \mathbb{P}[N_T^{(k)}]\mathbb{Q}\Big[\frac{1}{N_T^{(k)}}f(\xi_T)\Big]. \quad (3.6)$$

Dividing through by $\mathbb{P}(N_T\geq k)$ and applying Lemma 3.4.5 gives the result. □

3.4.4 The symmetry lemma

In this section we aim to show that $\mathbb{Q}^{k,T}$ satisfies a time-dependent Markov branching property, in that the descendants of any particle behave independently of the rest of the tree.

Lemma 3.4.6 (Symmetry lemma)**.** *Suppose that $v\in\mathcal{N}_t$ is carrying j marks at time t. Then, under $\mathbb{Q}^{k,T}$, the subtree generated by v after time t is independent of the rest of the system and behaves as if under $\mathbb{Q}^{j,T-t}$.*

Proof. Fix t,T and v. Let $\mathcal{H}$ be the σ-algebra generated by all the information except in the subtree generated by v after time t. Then it suffices to show that for $s\in(t,T]$ and $i\geq 0$,

$$\mathbb{Q}^{k,T}(\tau_v > s,\, L_v = i|\mathcal{H}) = \mathbb{Q}^{j,T-t}(\tau_\emptyset > s-t,\, L_\emptyset = i)$$

almost surely.

Recall that

$$g_{k,T} = \mathbb{1}_{\{\xi_T^i\neq\xi_T^j\ \forall i\neq j\}}\prod_{i=1}^k\prod_{v<\xi_T^i} L_v$$

and

$$\zeta_{k,T} = \frac{g_{k,T}}{\mathbb{P}[N_T^{(k)}]}.$$

Let I be the set of marks carried by v at time t, and let

$$\tilde{g} = \mathbb{1}_{\{\xi_T^i\neq\xi_T^j,\ \forall i\neq j,\, i,j\in I^c\}}\prod_{i\in I}\prod_{\xi_t^i\leq v<\xi_T^i} L_v$$

and

$$h = \mathbb{1}_{\{\xi^i_T \neq \xi^j_T \ \forall i \neq j,\, i,j \in I^c\}} \Big(\prod_{i \notin I} \prod_{v < \xi^i_T} L_v \Big) \prod_{i \in I} \prod_{v < \xi^i_t} L_v.$$

Note that h is $\mathcal{H}$-measurable and $g_{k,T} = \tilde{g}h$.

By Lemma **??**, $\mathbb{Q}^{k,T}$-almost surely,

$$\mathbb{Q}^{k,T}(\tau_v > s,\, L_v = i | \mathcal{H}) = \frac{1}{\mathbb{P}^k[\zeta_{k,T}|\mathcal{H}]} \mathbb{P}^k[\zeta_{k,T} \mathbb{1}_{\{\tau_v > s,\, L_v = i\}} | \mathcal{H}].$$

Cancelling factors of $\mathbb{P}^k[N^{(k)}_T]$ and using the fact that $g_{k,T} = \tilde{g}h$ where h is $\mathcal{H}$-measurable, we get

$$\mathbb{Q}^{k,T}(\tau_v > s,\, L_v = i | \mathcal{H}) = \frac{1}{h\mathbb{P}^k[\tilde{g}|\mathcal{H}]} h\mathbb{P}^k[\tilde{g}\mathbb{1}_{\{\tau_v > s,\, L_v = i\}} | \mathcal{H}] = \frac{\mathbb{P}^k[\tilde{g}\mathbb{1}_{\{\tau_v > s,\, L_v = i\}} | \mathcal{H}]}{\mathbb{P}^k[\tilde{g}|\mathcal{H}]}.$$

By the Markov branching property under $\mathbb{P}^k$, the behaviour of the subtree generated by v after time t is independent of the rest of the system and—on the event that v is carrying j marks at time t—behaves as if under $\mathbb{P}^j$. Thus a almost surely. Applying Lemma 3.4.1 establishes the result. □

3.4.5 Spine births under $\mathbb{Q}$

We already know from (3.5) and the discussion following it that particles that are not spines behave exactly as under $\mathbb{P}^k$: they branch at rate r and have offspring distribution L. The behaviour of the spine particles is more complicated.

Recall that $\tau_\emptyset$ is the first branching event, and ψ_1 is the time of the first spine splitting event, i.e.

$$\psi_1 = \inf\{t \geq 0 : \exists i, j \text{ with } \xi^i_t \neq \xi^j_t\}.$$

(Note that if the spines die without giving birth to any children, this counts as a splitting event.) By the symmetry lemma, in order to understand how the spines behave under $\mathbb{Q}$, it suffices to understand the distributions of $\tau_\emptyset$ and ψ_1.

Lemma 3.4.7. For any $t \in [0, T]$ and $k \geq 0$, we have

$$\mathbb{Q}^{k,T}(\tau_\phi > t) = \frac{\mathbb{P}^k[N^{(k)}_{T-t}]}{\mathbb{P}^k[N^{(k)}_T]} e^{-rt},$$

$$\mathbb{Q}^{k,T}(\psi_1 > t) = \frac{\mathbb{P}^k[N^{(k)}_{T-t}]}{\mathbb{P}^k[N^{(k)}_T]} e^{(m-1)rt},$$

and

$$\mathbb{Q}^{k,T}(\tau_\phi > t | \psi_1 > t) = e^{-mrt}.$$

Proof. For the first statement,

$$\mathbb{Q}(\tau_\phi > t) = \mathbb{P}[\zeta_{k,T}\mathbb{1}_{\{\tau_\emptyset > t\}}] = \frac{1}{\mathbb{P}[N^{(k)}_T]} \mathbb{P}[g_{k,T}\mathbb{1}_{\{\tau_\emptyset > t\}}].$$

By the Markov property and Lemma 3.4.1,

$$\mathbb{P}[g_{k,T}\mathbb{1}_{\{\tau_\emptyset > t\}}] = \mathbb{P}(\tau_\emptyset > t)\mathbb{P}[g_{k,T-t}] = e^{-rt}\mathbb{P}[N^{(k)}_{T-t}]$$

as required.

For the second statement,

$$\mathbb{Q}(\psi_1 > t) = \mathbb{P}[\zeta_{k,T}\mathbb{1}_{\{\psi_1>t\}}] = \frac{1}{\mathbb{P}[N_T^{(k)}]}\mathbb{P}[g_{k,T}\mathbb{1}_{\{\psi_1>t\}}],$$

and by the Markov property and Lemma 3.4.1,

$$\mathbb{P}[g_{k,T}\mathbb{1}_{\{\psi_1>t\}}] = \mathbb{P}\Big[\Big(\prod_{v<\xi_t^1} L_v^k\Big)\mathbb{1}_{\{\psi_1>t\}}\Big]\mathbb{P}[g_{k,T-t}] = \mathbb{P}\Big[\Big(\prod_{v<\xi_t^1} L_v^k\Big)\mathbb{1}_{\{\psi_1>t\}}\Big]\mathbb{P}[N_{T-t}^{(k)}].$$

Putting these two lines together we get

$$\mathbb{Q}(\psi_1 > t) = \frac{\mathbb{P}^k[N_{T-t}^{(k)}]}{\mathbb{P}^k[N_T^{(k)}]}\mathbb{P}\Big[\Big(\prod_{v<\xi_t^1} L_v^k\Big)\mathbb{1}_{\{\psi_1>t\}}\Big]. \tag{3.7}$$

Note that $\psi > t$ if and only if all k marks are following the same particle at time t (which must also be alive); thus

$$\mathbb{P}\Big[\Big(\prod_{v<\xi_t^1} L_v^k\Big)\mathbb{1}_{\{\psi_1>t\}}\Big] = \mathbb{P}\Big[\sum_{u\in\mathcal{N}_t}\Big(\prod_{v<u} L_v^k\Big)\mathbb{1}_{\{\xi_t^1=\ldots=\xi_t^k=u\}}\Big] = \mathbb{P}\Big[\sum_{u\in\mathcal{N}_t} 1\Big] = \mathbb{P}[N_t] = e^{(m-1)t}.$$

Substituting back into (3.7) gives the desired result.

The third statement follows easily from the first two. $\square$

The third part of Lemma 3.4.7 combined with the symmetry lemma (Lemma 3.4.6) tells us the following: given $\mathcal{G}_T^k$ (the information only about spine splitting events), under $\mathbb{Q}^{k,T}$ each spine gives birth to non-spine particles according to a Poisson process of rate mr, independently of everything else. In particular when there are n distinct spines alive, there are n independent Poisson point processes and the total rate at which non-spine particles are immigrated along the spines is nmr.

We call birth events that occur along the spines, but which do not occur at spine splitting events, *births off the spine*. The following lemma tells us the distribution of the number of children born at such events.

Lemma 3.4.8. For any $j \geq 0$, $k \geq 1$ and $0 \leq t < T$,

$$\mathbb{Q}^{k,T}(L_\emptyset = j | \tau_\emptyset = t,\, \psi_1 > t) = \frac{jp_j}{m}.$$

Proof. From the definition of $\mathbb{Q}$,

$$\begin{aligned}\mathbb{Q}(L_\emptyset = j|\tau_\emptyset = t,\, \psi_1 > t) &= \frac{\mathbb{P}[\zeta_{k,T}\mathbb{1}_{\{L_\emptyset=j\}}|\tau_\emptyset = t,\, \psi_1 > t]}{\mathbb{P}[\zeta_{k,T}|\tau_\emptyset = t,\, \psi_1 > t]}\\ &= \frac{\mathbb{P}[g_{k,T}\mathbb{1}_{\{L_\emptyset=j\}}|\tau_\emptyset = t,\, \psi_1 > t]}{\mathbb{P}[g_{k,T}|\tau_\emptyset = t,\, \psi_1 > t]}\\ &= \frac{\mathbb{P}[g_{k,T}\mathbb{1}_{\{L_\emptyset=j,\,\psi_1>t\}}|\tau_\emptyset = t]}{\mathbb{P}[g_{k,T}\mathbb{1}_{\{\psi_1>t\}}|\tau_\emptyset = t]}.\end{aligned}$$

By the Markov property, for any i,

$$\mathbb{P}[g_{k,T}\mathbb{1}_{\{L_\emptyset=i,\,\psi_1>t\}}|\tau_\emptyset = t] = p_i i^k\frac{1}{i^{k-1}}\mathbb{P}[g_{k,T-t}] = ip_i\mathbb{P}[g_{k,T-t}].$$

Thus

$$\mathbb{Q}(L_\emptyset = j | \tau_\emptyset = t,\, \psi_1 > t) = \frac{jp_j \mathbb{P}[g_{k,T-t}]}{\sum_i ip_i \mathbb{P}[g_{k,T-t}]} = \frac{jp_j}{m}. \qquad \square$$

A random variable that takes the value j with probability jp_j/m for each j is said to be *size-biased* (relative to L). Lemma 3.4.8 then tells us (in conjunction with the symmetry lemma) that births off any spine are always size-biased, no matter how many marks are following that particular spine. (The number of marks therefore only affects spine splitting events.)

Our next result gives us a bound on the number of children born at any branching event under $\mathbb{Q}$; it is only a rough bound, but we will use it to control the number of residue particles.

Lemma 3.4.9. There exists an auxiliary random variable $\hat{L}_k$, depending only on k, such that

$$\mathbb{Q}^{k,T}(L_\emptyset \geq j) \leq P(\hat{L}_k \geq j)$$

for all $j \geq 0$ and $T > 0$.

Again we stress that in conjunction with the symmetry lemma, this tells us about *any* branching event under $\mathbb{Q}$, not just the first one.

Proof. Let n_t be the number of distinct spine particles at time t. Then, proceeding similarly to Lemma 3.4.8,

$$\begin{aligned}\mathbb{Q}^{k,T}(L_\emptyset = j | \tau_\emptyset = t,\, n_t = l) &= \frac{\mathbb{P}[\zeta_{k,T} \mathbb{1}_{\{L_\emptyset = j,\, n_t = l\}} | \tau_\emptyset = t]}{\mathbb{P}[\zeta_{k,T} \mathbb{1}_{\{n_t = l\}} | \tau_\emptyset = t]} \\ &= \frac{\mathbb{P}[g_{k,T} \mathbb{1}_{\{L_\emptyset = j,\, n_t = l\}} | \tau_\emptyset = t]}{\mathbb{P}[g_{k,T} \mathbb{1}_{\{n_t = l\}} | \tau_\emptyset = t]}.\end{aligned}$$

Then by the Markov property,

$$\mathbb{P}[g_{k,T} \mathbb{1}_{\{L_\emptyset = j,\, n_t = l\}} | \tau_\emptyset = t] = p_j j^k \binom{j}{l} \sum_{\substack{a_1,\ldots,a_l \in \{1,\ldots,k\}: \\ a_1 + \ldots + a_l = k}} \frac{k!}{a_1! \ldots a_l!} \frac{1}{j^k} \prod_{i=1}^{l} \mathbb{P}[g_{a_i, T-t}],$$

where p_j is the probability that the number of children at the first branching event is j; j^k is the first factor in the product in the definition of $g_{k,T}$; $\binom{j}{l}$ is the number of ways of choosing l particles to be spines at time t; the sum of terms involving $a_1, \ldots, a_l$ is the number of ways of assigning the k marks to the chosen l spine particles; $1/j^k$ is the probability of the k marks all following their assigned spine particle; and $\prod_{i=1}^{l} \mathbb{P}[g_{a_i,T-t}]$ is the remaining contribution from $g_{k,T}$ after t. Thus

$$\begin{aligned}\mathbb{Q}^{k,T}(L_\emptyset = j | \tau_\emptyset = t,\, n_t = l) &= \frac{p_j j^k \binom{j}{l} \sum_{a_1,\ldots,a_l} \frac{k!}{a_1! \ldots a_l!} \frac{1}{j^k} \prod_{i=1}^{l} \mathbb{P}[g_{a_i,T-t}]}{\sum_b p_b b^k \binom{b}{l} \sum_{a_1,\ldots,a_l} \frac{k!}{a_1! \ldots a_l!} \frac{1}{b^k} \prod_{i=1}^{l} \mathbb{P}[g_{a_i,T-t}]} \\ &= \frac{p_j j^{(l)}}{\mathbb{P}[L^{(l)}]}.\end{aligned}$$

Since this does not depend on t, we get

$$\mathbb{Q}^{k,T}(L_\emptyset = j | n_{\tau_\phi} = l) = \frac{p_j j^{(l)}}{\mathbb{P}[L^{(l)}]}.$$

Letting $\hat{L}(l)$ be an auxiliary random variable with $P(\hat{L}(l) = j) = p_j j^{(l)} / \mathbb{P}[L^{(l)}]$, and $\hat{L}_k = \max_{l=1,\ldots,k} \hat{L}(l)$, we have

$$\mathbb{Q}^{k,T}(L_\emptyset \geq j) \leq P(\hat{L}_k \geq j). \qquad \square$$

3.4.6 Campbell's formula

One of the key elements that we need to carry out our calculations will be a version of Campbell's formula. Let $\tilde{N}_t$ be the number of *ordinary* particles alive at time t—that is, they are not spines, and did not split from spines at spine splitting events. Recall that we also defined n_t to be the number of distinct spines alive at time t.

We write $F(\theta, t) = \mathbb{P}[\theta^{N_t}]$ and $u(\theta) = \mathbb{P}[\theta^L] - \theta$. These functions satisfy the Kolmogorov forwards and backwards equations

$$\frac{\partial}{\partial t} F(\theta, t) = ru(\theta) \frac{\partial}{\partial \theta} F(\theta, t) \tag{3.8}$$

and

$$\frac{\partial}{\partial t} F(\theta, t) = ru(F(\theta, t)); \tag{3.9}$$

see [5, Chapter III, Section 3]. Our main aim is to show the following.

Proposition 3.4.10. For any $z \geq 0$,

$$\mathbb{Q}^{k,T}[e^{-z\tilde{N}_T} | \mathcal{G}_T^k] = \prod_{i=0}^{k-1} \Big(e^{-r(m-1)(T-\psi_i)} \frac{u(F(e^{-z}, T - \psi_i))}{u(e^{-z})} \Big)$$

$\mathbb{Q}^{k,T}$-almost surely.

Notice in particular that the right-hand side depends only on the values of the split times of the spines, not any of the other information in $\mathcal{G}_T^k$ (for example the topological information about the tree). This—used in conjunction with Proposition 3.4.2—is a large part of the reason that the split times of our k uniformly chosen particles are (asymptotically) independent of the topological information in the induced tree.

The main step in proving Proposition 3.4.10 comes from the next lemma.

Lemma 3.4.11. For any $z \geq 0$,

$$\mathbb{Q}[e^{-z\tilde{N}_T} | \mathcal{G}_T^k] = \prod_{i=0}^{k-1} \exp\Big(-r(m-1)(T-\psi_i) + r \int_0^{T-\psi_i} u'(\mathbb{P}[e^{-zN_s}]) \, \mathrm{d}s \Big).$$

$\mathbb{Q}^{k,T}$-almost surely.

Proof. Let Λ_T be the total number of birth events off the spines (i.e. births along spines that are not spine splitting events) before time T. Recall (from Lemma 3.4.7 and the symmetry lemma) that under $\mathbb{Q}^{k,T}$ each spine gives birth to non-spine particles according to a Poisson process of rate rm, independently of everything else. Thus at any time $s \in [0, T]$, the total rate at which spine particles give birth to non-spine particles is rmn_s. Besides, such births are size biased (by Lemma 3.4.8 and the symmetry lemma). Finally, once a particle is born off the spines, it generates a tree that behaves exactly as under $\mathbb{P}$ (see (3.5) and the discussion that follows).

Thus, letting $\lambda_T = \int_0^T n_s \mathrm{d}s$,

$$\mathbb{Q}[e^{-z\tilde{N}_T} | \mathcal{G}_T^k] = \sum_{j=0}^{\infty} \mathbb{Q}(\Lambda_T = j) \Big(\int_0^T \sum_{i=1}^{\infty} \frac{ip_i^{(T)}}{m} \mathbb{P}[e^{-zN_{T-s}}]^{i-1} \frac{n_s}{\lambda_T} \mathrm{d}s \Big)^j.$$

Since $\mathbb{Q}(\Lambda_T = j) = e^{-rm\lambda_T}(rm\lambda_T)^j/j!$, we get

$$\mathbb{Q}[e^{-z\tilde{N}_T}|\mathcal{G}_T^k] = e^{-rm\lambda_T}\sum_{j=0}^{\infty}\frac{1}{j!}\Big(r\int_0^T\sum_{i=1}^{\infty} ip_i^{(T)}\mathbb{P}[e^{-zN_{T-s}}]^{i-1}n_s\mathrm{d}s\Big)^j.$$

Note that

$$\sum_{i=1}^{\infty} ip_i^{(T)}\theta^{i-1} = \frac{\mathrm{d}}{\mathrm{d}\theta}\sum_{i=1}^{\infty}p_i^{(T)}\theta^i = u'(\theta)+1.$$

Therefore

$$\mathbb{Q}[e^{-z\tilde{N}_T}|\mathcal{G}_T^k] = \exp\Big(-r(m-1)\lambda_T + r\int_0^T u'(\mathbb{P}[e^{-zN_{T-s}}])n_s\mathrm{d}s\Big).$$

Now, we know that between times ψ_{i-1} and ψ_i we have exactly i distinct spine particles. Thus

$$\mathbb{Q}[e^{-z\tilde{N}_T}|\mathcal{G}_T^k] = \prod_{i=0}^{k-1}\exp\Big(-r(m-1)(T-\psi_i) + r\int_{\psi_i}^T u'(\mathbb{P}[e^{-zN_{T-s}}])\,\mathrm{d}s\Big)$$

and the result follows. □

Proof of Proposition 3.4.10. Recalling (3.9) that $F(\theta, s)$ satisfies the backwards equation

$$\frac{\partial}{\partial s}F(\theta,s) = ru(F(\theta,s)),$$

by making the substitution $t = F(\theta, s)$ we see that

$$r\int_a^b u'(F(\theta,s))\mathrm{d}s = r\int_{F(\theta,a)}^{F(\theta,b)}\frac{u'(t)}{ru(t)}\mathrm{d}t = \log\Big(\frac{u(F(\theta,b))}{u(F(\theta,a))}\Big).$$

Applying this to Lemma 3.4.11, we have

$$\mathbb{Q}[e^{-z\tilde{N}_T}|\mathcal{G}_T^k] = \prod_{i=0}^{k-1}\Big(e^{-r(m-1)(T-\psi_i)}\frac{u(F(e^{-z},T-\psi_i))}{u(F(e^{-z},0))}\Big).$$

Noting that $F(e^{-z}, 0) = e^{-z}$ gives the result. □

3.5. Birth-death processes

In this section we aim to prove the results from Section 3.2.1. Recall the setup: fix $a \geq 0$ and $b > 0$, and suppose that $r = \alpha+\beta$, $p_0 = \alpha/(\alpha+\beta)$ and $p_2 = \beta/(\alpha+\beta)$, with $p_j = 0$ for $j \neq 0, 2$. This is known as a birth-death process with birth rate β and death rate α. Since all particles have either 0 or 2 children, and under $\mathbb{Q}$ the spines cannot have 0 children, they must always have 2 children. This simplifies the picture considerably.

3.5.1 Elementary calculations with generating functions

Suppose first that we are in the non-critical case $\alpha \neq \beta$. It is easy to calculate the moment generating function under $\mathbb{P}$ for a birth-death process (see [5, Chapter III, Section 5]): for $\alpha \neq \beta$ and $\theta \in (0, 1)$,

$$F(\theta,t) := \mathbb{P}[\theta^{N_t}] = \frac{\alpha(1-\theta)e^{(\beta-\alpha)t}+\beta\theta-\alpha}{\beta(1-\theta)e^{(\beta-\alpha)t}+\beta\theta-\alpha}.$$

We then see that

$$\mathbb{P}(N_t = 0) = \lim_{\theta\downarrow 0} F(0,t) = \frac{\alpha e^{(\beta-\alpha)t} - \alpha}{\beta e^{(\beta-\alpha)t} - \alpha}.$$

Writing

$$p_t = \mathbb{P}(N_t = 0) = \frac{\alpha e^{(\beta-\alpha)t} - \alpha}{\beta e^{(\beta-\alpha)t} - \alpha}, \qquad 1 - p_t = \frac{(\beta-\alpha)e^{(\beta-\alpha)t}}{\beta e^{(\beta-\alpha)t} - \alpha}$$

and

$$q_t = \frac{\beta e^{(\beta-\alpha)t} - \beta}{\beta e^{(\beta-\alpha)t} - \alpha}, \qquad 1 - q_t = \frac{\beta - \alpha}{\beta e^{(\beta-\alpha)t} - \alpha},$$

we get

$$F(\theta,t) = p_t + (1-p_t)\frac{(1-q_t)\theta}{1-q_t\theta} = p_t + \frac{(1-p_t)(1-q_t)}{q_t}\Big(\frac{1}{1-q_t\theta} - 1\Big).$$

From this we see that

$$F(\theta,t) = p_t + (1-p_t)(1-q_t)\sum_{j=1}^{\infty}\theta^j q_t^{j-1}$$

and

$$\frac{\partial^k F(\theta,t)}{\partial\theta^k} = \frac{(1-p_t)(1-q_t)}{q_t}\frac{q_t^k k!}{(1-q_t\theta)^{k+1}}.$$

Therefore

$$\mathbb{P}(N_t = j) = (1-p_t)(1-q_t)q_t^{j-1} \quad \text{for } j \geq 1,$$

so

$$\mathbb{P}(N_t \geq k) = (1-p_t)(1-q_t)\sum_{j=k}^{\infty} q_t^{j-1} = (1-p_t)q_t^{k-1} = \frac{(\beta-\alpha)e^{(\beta-\alpha)t}\beta^{k-1}(e^{(\beta-\alpha)t}-1)^{k-1}}{(\beta e^{(\beta-\alpha)t} - \alpha)^k}.$$

Also, since $\mathbb{P}[N_t^{(k)}] = \lim_{\theta\uparrow 1}\frac{\partial^k F(\theta,t)}{\partial\theta^k}$,

$$\mathbb{P}[N_t^{(k)}] = \frac{(1-p_t)(1-q_t)}{q_t}\frac{q_t^k k!}{(1-q_t)^{k+1}} = k!\Big(\frac{\beta}{\beta-\alpha}\Big)^{k-1} e^{(\beta-\alpha)t}(e^{(\beta-\alpha)t}-1)^{k-1}. \tag{3.10}$$

Thus

$$\frac{\mathbb{P}[N_t^{(k)}]}{\mathbb{P}(N_t \geq k)} = \frac{k!}{(\beta-\alpha)^k}(\beta e^{(\beta-\alpha)t} - \alpha)^k \tag{3.11}$$

and

$$\frac{\mathbb{P}\big[N_{T-t}^{(k)}\big]}{\mathbb{P}\big[N_T^{(k)}\big]} = e^{-(\beta-\alpha)t}\Big(\frac{e^{(\beta-\alpha)(T-t)} - 1}{e^{(\beta-\alpha)T} - 1}\Big)^{k-1}. \tag{3.12}$$

Finally, writing

$$F(\theta,t) = \frac{\alpha}{\beta} + \frac{(\beta-\alpha)\theta - \alpha(\beta-\alpha)/\beta}{\beta(1-\theta)e^{(\beta-\alpha)t} + \beta\theta - \alpha},$$

we see that

$$\frac{\partial F(\theta,t)}{\partial t} = \frac{(\beta-\alpha)^2(\beta\theta-\alpha)(1-\theta)e^{(\beta-\alpha)t}}{(\beta(1-\theta)e^{(\beta-\alpha)t} + \beta\theta - \alpha)^2}. \tag{3.13}$$

In the critical case $\alpha = \beta$, similar calculations give

$$F(\theta,t) := \mathbb{P}[\theta^{N_t}] = \frac{(1-\theta)\beta t + \theta}{(1-\theta)\beta t + 1}, \tag{3.14}$$

$$\mathbb{P}[N_t^{(k)}] = \lim_{\theta\uparrow 1} \frac{\partial^k F(\theta,t)}{\partial\theta^k} = k!(\beta t)^{k-1}, \tag{3.15}$$

$$\frac{\mathbb{P}[N_t^{(k)}]}{\mathbb{P}(N_t \geq k)} = k!(\beta t+1)^k \tag{3.16}$$

and

$$\frac{\partial F(\theta,t)}{\partial t} = \frac{\partial}{\partial t}\Big(1+\frac{\theta-1}{(1-\theta)\beta t+1}\Big) = \frac{(1-\theta)^2\beta}{((1-\theta)\beta t+1)^2}. \tag{3.17}$$

3.5.2 Split time densities

Recall that $\mathcal{H}'$ is the σ-algebra that contains information about which marks follow which spines, but does not know anything about the spine split times.

Lemma 3.5.1. Under $\mathbb{Q}^{k,T}$, the spine split times $\psi_1,\ldots,\psi_{k-1}$ are independent of $\mathcal{H}'$ and have a joint probability density function

$$f_k^{\mathbb{Q}}(s_1,\ldots,s_{k-1}) = \begin{cases} (k-1)!\Big(\frac{\beta-\alpha}{e^{(\beta-\alpha)T}-1}\Big)^{k-1}\prod_{i=1}^{k-1} e^{(\beta-\alpha)(T-s_i)} & \text{if } \alpha\neq\beta \\ (k-1)!/T^{k-1} & \text{if } \alpha=\beta \end{cases}.$$

Proof. We do the calculation in the non-critical case $\alpha \neq \beta$. The proof in the critical case is identical.

Recall from Lemma 3.4.7 that

$$\mathbb{Q}^{k,T}(\psi_1 > s_1) = \frac{\mathbb{P}[N_{T-s_1}^{(k)}]}{\mathbb{P}[N_T^{(k)}]}e^{(m-1)rs_1} = \frac{\mathbb{P}[N_{T-s_1}^{(k)}]}{\mathbb{P}[N_T^{(k)}]}e^{(\beta-\alpha)s_1}.$$

Then (3.12) gives

$$\mathbb{Q}^{k,T}(\psi_1 > s_1) = e^{-(\beta-\alpha)s_1}\Big(\frac{e^{(\beta-\alpha)(T-s_1)}-1}{e^{(\beta-\alpha)T}-1}\Big)^{k-1}e^{(\beta-\alpha)s_1} = \Big(\frac{e^{(\beta-\alpha)(T-s_1)}-1}{e^{(\beta-\alpha)T}-1}\Big)^{k-1},$$

so ψ_1 has density

$$(k-1)(\beta-\alpha)e^{(\beta-\alpha)(T-s_1)}\frac{(e^{(\beta-\alpha)(T-s_1)}-1)^{k-2}}{(e^{(\beta-\alpha)T}-1)^{k-1}}.$$

For $i=2,\ldots,k-1$, between times ψ_{i-1} and ψ_i we have exactly i particles carrying marks. Let A_i be the event that the first of these is carrying a_1 marks, the second a_2, and so on. Let $\psi_i^{(j)}$ be the time at which the marks following the jth of these particles split. By the symmetry lemma, given $\psi_{i-1}=s_{i-1}$, these times are independent with

$$\mathbb{Q}^{k,T}(\psi_i^{(j)} > s_i|\psi_{i-1}=s_{i-1},\,A_i) = \mathbb{Q}^{a_j,T-s_{i-1}}(\psi_1 > s_i-s_{i-1}) = \Big(\frac{e^{(\beta-\alpha)(T-s_i)}-1}{e^{(\beta-\alpha)(T-s_{i-1})}-1}\Big)^{a_j-1}.$$

Then, since the event $\{\psi_i > s_i\} = \bigcap_j\{\psi_i^{(j)} > s_i\}$,

$$\mathbb{Q}^{k,T}(\psi_i > s_i|\psi_{i-1}=s_{i-1},\,A_i) = \prod_{j=1}^{i}\Big(\frac{e^{(\beta-\alpha)(T-s_i)}-1}{e^{(\beta-\alpha)(T-s_{i-1})}-1}\Big)^{a_j-1}.$$

Since $\sum_{j=1}^{i}(a_j-1)=k-i$, we get

$$\mathbb{Q}^{k,T}(\psi_i > s_i|\psi_{i-1}=s_{i-1},\, A_i) = \Big(\frac{e^{(\beta-\alpha)(T-s_i)}-1}{e^{(\beta-\alpha)(T-s_{i-1})}-1}\Big)^{k-i}.$$

This does not depend on $a_1,\dots,a_i$, so ψ_i is independent of $\mathcal{H}'$, and summing over the possible values we obtain

$$\mathbb{Q}^{k,T}(\psi_i > s_i \,|\, \psi_{i-1}=s_{i-1}) = \Big(\frac{e^{(\beta-\alpha)(T-s_i)}-1}{e^{(\beta-\alpha)(T-s_{i-1})}-1}\Big)^{k-i}.$$

Differentiating gives

$$f_k^{\mathbb{Q}}(s_1,\dots,s_{k-1}) = (k-1)!(\beta-\alpha)^{k-1}\prod_{i=1}^{k-1} e^{(\beta-\alpha)(T-s_i)}\frac{(e^{(\beta-\alpha)(T-s_i)}-1)^{k-i-1}}{(e^{(\beta-\alpha)(T-s_{i-1}}-1)^{k-i}}.$$

The product telescopes to give the answer. □

Proposition 3.5.2. Let $s_0=0$. The vector $(\mathcal{S}_1^k(T),\dots,\mathcal{S}_{k-1}^k(T))$ of ordered split times under $\mathbb{P}$ is independent of $\mathcal{H}$ and has a joint density $f_k^T(s_1,\dots,s_{k-1})$ equalling

$$\frac{k!(\beta e^{(\beta-\alpha)T}-\alpha)^k(\beta-\alpha)^{2k-1}}{(e^{(\beta-\alpha)T}-1)^{k-1}e^{(\beta-\alpha)T}}\int_0^1(1-y)^{k-1}\prod_{j=0}^{k-1}\frac{e^{(\beta-\alpha)(T-s_j)}}{(\beta(1-y)e^{(\beta-\alpha)(T-s_j)}+\beta y-\alpha)^2}\mathrm{d}y$$

if $\alpha\neq\beta$, and

$$\frac{k!(\beta T+1)^k}{T^{k-1}}\int_0^1(1-y)^{k-1}\prod_{j=0}^{k-1}\frac{1}{(\beta(1-y)(T-s_j)+1)^2}\mathrm{d}y$$

if $\alpha=\beta$.

Proof. Again we give the proof in the non-critical case $\alpha\neq\beta$. The critical case is identical. We start with Proposition 3.4.2, which tells us that for any measurable functional F,

$$\mathbb{P}\Big[\frac{1}{N_T^{(k)}}\sum_{u\in\mathcal{N}_T^{(k)}}F(u)\,\Big|\,N_T\geq k\Big] = \frac{\mathbb{P}[N_T^{(k)}]}{\mathbb{P}(N_T\geq k)(k-1)!}\int_0^\infty(e^z-1)^{k-1}\mathbb{Q}^{k,T}\Big[e^{-zN_T}F(\xi_T)\Big]\,\mathrm{d}z. \tag{3.18}$$

The independence of the spine split times and $\mathcal{H}'$ under $\mathbb{Q}^{k,T}$ (established in Lemma 3.5.1), together with (3.18) and Proposition 3.4.10, imply that the split times under $\mathbb{P}$ are independent of $\mathcal{H}$.

Returning to (3.18) again, we get that in particular

$$\begin{aligned}&f_k^T(s_1,\dots,s_{k-1})\\&=\frac{\mathbb{P}[N_T^{(k)}]}{\mathbb{P}(N_T\geq k)(k-1)!}\int_0^\infty(e^z-1)^{k-1}f_k^{\mathbb{Q}}(s_1,\dots,s_{k-1})\mathbb{Q}[e^{-zN_T}|\psi_1=s_1,\dots,\psi_{k-1}=s_{k-1}]\,\mathrm{d}z.\end{aligned}$$

However we also know from Proposition 3.4.10 that

$$\mathbb{Q}[e^{-z\tilde{N}_T}|\psi_1=s_1,\dots,\psi_{k-1}=s_{k-1}] = \prod_{i=0}^{k-1}\Big(e^{-r(m-1)(T-s_i)}\frac{u(F(e^{-z},T-s_i))}{u(e^{-z})}\Big)$$

where $s_0 = 0$, $F(\theta, t) = \mathbb{P}[\theta^{N_t}]$ and $u(\theta) = \mathbb{P}[\theta^L] - \theta$. Of course since all births are binary, all particles are either spines or ordinary; so since there are k spines at time T almost surely under $\mathbb{Q}$, $N_T = \tilde{N}_T + k$. Thus, by (3.9) and (3.13),

$$\mathbb{Q}[e^{-zN_T}|\psi_1 = s_1, \ldots, \psi_{k-1} = s_{k-1}] = e^{-zk} \prod_{i=0}^{k-1} \Big(\frac{\beta - \alpha}{\beta(1 - e^{-z})e^{(\beta-\alpha)(T-s_i)} + \beta e^{-z} - \alpha} \Big)^2.$$

Plugging this into our formula for $f_k^T(s_1, \ldots, s_{k-1})$ above gives

$$f_k^T(s_1, \ldots, s_{k-1}) = \frac{\mathbb{P}[N_T^{(k)}]}{\mathbb{P}(N_T \geq k)(k-1)!} \int_0^\infty e^{-z}(1 - e^{-z})^{k-1} f_k^{\mathbb{Q}}(s_1, \ldots, s_{k-1}) \cdot \prod_{i=0}^{k-1} \frac{(\beta - \alpha)^2}{(\beta(1 - e^{-z})e^{(\beta-\alpha)(T-s_i)} + \beta e^{-z} - \alpha)^2} \, \mathrm{d}z.$$

By (3.11) and Lemma 3.5.1, this becomes

$$\frac{k!(\beta e^{(\beta-\alpha)T} - \alpha)^k (\beta - \alpha)^{2k-1}}{e^{(\beta-\alpha)T}(e^{(\beta-\alpha)T} - 1)^{k-1}} \int_0^\infty e^{-z}(1-e^{-z})^{k-1} \prod_{i=0}^{k-1} \frac{e^{(\beta-\alpha)(T-s_i)}}{(\beta(1 - e^{-z})e^{(\beta-\alpha)(T-s_i)} + \beta e^{-z} - \alpha)^2} \mathrm{d}y.$$

Making the substitution $y = e^{-z}$ gives the result. □

3.5.3 Describing the partition process

We recall now the partition $Z_0, Z_1, \ldots$ which contained the information about the marks following each of the distinct spine particles, without the information about the split times.

Lemma 3.5.3. The partition $Z_0, Z_1, \ldots$ has the following distribution under $\mathbb{Q}_T^{k,T}$:

- If Z_i consists of $i+1$ blocks of sizes $a_1, \ldots, a_{i+1}$, then the jth block will split next with probability $\frac{a_j - 1}{k-i-1}$ for each $j = 1, \ldots, i+1$.
- When a block of size a splits, it splits into two new blocks, and the probability that these blocks have sizes l and $a - l$ is $\frac{1}{a-1}$ for each $l = 1, \ldots, a-1$.

Proof. Suppose that we are given $\psi_i = s$. For the first part, by the symmetry lemma, the probability that the jth block splits next is

$$\int_0^{T-s} \mathbb{Q}^{a_j, T-s}(\psi_1 \in \mathrm{d}t) \prod_{l \neq j} \mathbb{Q}^{a_l, T-s}(\psi_1 > t)$$

which by Lemma 3.4.7 equals

$$\int_0^{T-s} \Big(-\frac{\mathrm{d}}{\mathrm{d}t} \Big(\frac{\mathbb{P}[N_{T-s-t}^{(a_j)}]}{\mathbb{P}[N_{T-s}^{(a_j)}]} e^{(m-1)rt} \Big) \Big) \prod_{l \neq j} \frac{\mathbb{P}[N_{T-s-t}^{(a_l)}]}{\mathbb{P}[N_{T-s}^{(a_l)}]} e^{(m-1)rt} \, \mathrm{d}t.$$

If $\alpha \neq \beta$, then applying (3.12), the above becomes

$$\int_0^{T-s} \Big(-\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{e^{(\beta-\alpha)(T-s-t)}-1}{e^{(\beta-\alpha)(T-s)}-1}\Big)^{a_j-1}\Big) \prod_{l\neq j}\Big(\frac{e^{(\beta-\alpha)(T-s-t)}-1}{e^{(\beta-\alpha)(T-s)}-1}\Big)^{a_l-1}\mathrm{d}t$$
$$= (a_j-1)(\beta-\alpha)\int_0^{T-s} e^{(\beta-\alpha)(T-s-t)}\frac{(e^{(\beta-\alpha)(T-s-t)}-1)^{a_j-2}}{(e^{(\beta-\alpha)(T-s)}-1)^{a_j-1}}\prod_{l\neq j}\Big(\frac{e^{(\beta-\alpha)(T-s-t)}-1}{e^{(\beta-\alpha)(T-s)}-1}\Big)^{a_l-1}\mathrm{d}t$$
$$= (a_j-1)(\beta-\alpha)\int_0^{T-s} \frac{e^{(\beta-\alpha)(T-s-t)}}{e^{(\beta-\alpha)(T-s-t)}-1}\Big(\frac{e^{(\beta-\alpha)(T-s-t)}-1}{e^{(\beta-\alpha)(T-s)}-1}\Big)^{k-i-1}\mathrm{d}t.$$

Since the integrand does not depend on a_j, and we know the sum of the above quantity over $j = 1, \ldots, i+1$ must equal 1 (since one of the blocks must split first), we get

$$(\beta-\alpha)\int_0^{T-s} \frac{e^{(\beta-\alpha)(T-s-t)}}{e^{(\beta-\alpha)(T-s-t)}-1}\Big(\frac{e^{(\beta-\alpha)(T-s-t)}-1}{e^{(\beta-\alpha)(T-s)}-1}\Big)^{k-i-1}\mathrm{d}t = \frac{1}{k-i-1}$$

and therefore the probability that the jth block splits next equals $\frac{a_j-1}{k-i-1}$ as claimed. If $\alpha = \beta$ then applying (3.15) in place of (3.12) gives the same result.

For the second part, let ρ_t^1 be the number of marks following the first spine particle at time t. From the definition of $\mathbb{Q}^{k,T}$,

$$\mathbb{Q}^{k,T}(\rho_t^1 = i \,|\, \tau_\emptyset = t) = \frac{\mathbb{P}[g_{k,T}\mathbb{1}_{\{\rho_t^1=i\}} \,|\, \tau_\emptyset = t]}{\mathbb{P}[g_{k,T} \,|\, \tau_\emptyset = t]}.$$

By the Markov property, since each mark chooses uniformly from amongst the children available,

$$\mathbb{P}[g_{k,T}\mathbb{1}_{\{\rho_t^1=i\}} \,|\, \tau_\emptyset = t] = \frac{\beta}{\beta+\alpha}\binom{k}{i}\frac{1}{2^k}\mathbb{P}[g_{i,T-t}]\mathbb{P}[g_{k-i,T-t}].$$

Lemma 3.4.1 tells us that $\mathbb{P}[g_{j,s}] = \mathbb{P}[N_s^{(j)}]$ for any j and s, so

$$\mathbb{P}[g_{k,T}\mathbb{1}_{\{\rho_t^1=i\}} \,|\, \tau_\emptyset = t] = \frac{\beta}{\beta+\alpha}\binom{k}{i}\frac{1}{2^k}\mathbb{P}[N_{T-t}^{(i)}]\mathbb{P}[N_{T-t}^{(k-i)}].$$

If $\alpha \neq \beta$, then applying (3.10) gives

$$\mathbb{P}[g_{k,T}\mathbb{1}_{\{\rho_t^1=i\}} \,|\, \tau_\emptyset = t] = \frac{\beta}{\beta+\alpha}\binom{k}{i}\frac{1}{2^k}i!(k-i)!\Big(\frac{\beta}{\beta-\alpha}\Big)^{k-2}e^{(\beta-\alpha)(T-t)}(e^{(\beta-\alpha)(T-t)}-1)^{k-2}$$
$$= \frac{\beta}{\beta+\alpha}\frac{k!}{2^k}\Big(\frac{\beta}{\beta-\alpha}\Big)^{k-2}e^{(\beta-\alpha)(T-t)}(e^{(\beta-\alpha)(T-t)}-1)^{k-2}.$$

Since this does not depend upon i, we deduce that the distribution of ρ_t^1 under $\mathbb{Q}^{k,T}$ must be uniform. The case $\alpha = \beta$ is the same but using (3.15) in place of (3.10). The result now follows from the symmetry lemma. $\square$

3.5.4 Proofs of Theorems 3.2.1 and 3.2.2: explicit distribution functions for unordered split times

We now have all the ingredients to prove our theorem on the distribution of the split times. We begin with the non-critical case.

Proof of Theorem 3.2.1. By Proposition 3.5.2, the *ordered* split times are independent of $\mathcal{H}$ and have density

$$f_k^T(s_1,\ldots,s_{k-1}) = \frac{k!(\beta E_0-\alpha)^k(\beta-\alpha)^{2k-1}}{(E_0-1)^{k-1}E_0}\int_0^1 (1-y)^{k-1}\prod_{j=0}^{k-1}\frac{e^{(\beta-\alpha)(T-s_j)}}{(\beta(1-y)e^{(\beta-\alpha)(T-s_j)}+\beta y-\alpha)^2}\mathrm{d}y$$

for any $0\le s_1\le\ldots\le s_{k-1}\le 1$, where $s_0=0$. Therefore (see Lemma 3.7.3) the *unordered* split times are independent of $\mathcal{H}$ and have density

$$\tilde{f}_k^T(s_1,\ldots,s_{k-1}) = \frac{k(\beta E_0-\alpha)^k(\beta-\alpha)^{2k-1}}{(E_0-1)^{k-1}E_0}\int_0^1 (1-y)^{k-1}\prod_{j=0}^{k-1}\frac{e^{(\beta-\alpha)(T-s_j)}}{(\beta(1-y)e^{(\beta-\alpha)(T-s_j)}+\beta y-\alpha)^2}\mathrm{d}y.$$

Using Lemma 3.7.2 to integrate over s_j for each $j=1,\ldots,k-1$, we get

$$\begin{aligned}&\mathbb{P}(\tilde{\mathcal{S}}_1\ge s_1,\ldots,\tilde{\mathcal{S}}_{k-1}\ge s_{k-1}|N_T\ge k)\\ &=\frac{k(\beta E_0-\alpha)^k(\beta-\alpha)}{(E_0-1)^{k-1}E_0}\int_0^1(1-y)^{k-1}\left(\prod_{j=1}^{k-1}\frac{E_j-1}{\beta(1-y)E_j+\beta y-\alpha}\right)\frac{E_0}{(\beta(1-y)E_0+\beta y-\alpha)^2}\mathrm{d}y.\end{aligned}$$

Substituting $\theta=1-y$ and simplifying,

$$\begin{aligned}&\mathbb{P}(\tilde{\mathcal{S}}_1\ge s_1,\ldots,\tilde{\mathcal{S}}_{k-1}\ge s_{k-1}|N_T\ge k)\\ &\quad=\frac{\beta k(E_0-\alpha/\beta)^k(\beta-\alpha)}{(E_0-1)^{k-1}E_0}\int_0^1\left(\prod_{j=1}^{k-1}\frac{\theta(E_j-1)}{\theta E_j+1-\theta-\alpha/\beta}\right)\frac{E_0}{(\beta-\alpha+\beta\theta(E_0-1))^2}\mathrm{d}\theta\\ &\quad=\frac{\beta k(E_0-\alpha/\beta)^k}{(E_0-1)^{k-1}(\beta-\alpha)}\int_0^1\left(\prod_{j=1}^{k-1}\left(1-\frac{1}{1+\theta\frac{\beta}{\beta-\alpha}(E_j-1)}\right)\right)\frac{1}{(1+\theta\frac{\beta}{\beta-\alpha}(E_0-1))^2}\mathrm{d}\theta.\end{aligned}$$

We can now apply the second part of Lemma 3.7.1, with $e_j=\frac{\beta}{\beta-\alpha}(E_j-1)$ which gives

$$\begin{aligned}&\mathbb{P}(\tilde{\mathcal{S}}_1\ge s_1,\ldots,\tilde{\mathcal{S}}_{k-1}\ge s_{k-1}|N_T\ge k)\\ &\quad=\frac{\beta k(E_0-\alpha/\beta)^k}{(E_0-1)^{k-1}(\beta-\alpha)}\left[\frac{1}{1+e_0}\prod_{i=1}^{k-1}\frac{e_i}{e_i-e_0}+\sum_{j=1}^{k-1}\frac{e_j}{(e_j-e_0)^2}\left(\prod_{\substack{i=1\\ i\neq j}}^{k-1}\frac{e_i}{e_i-e_j}\right)\log\left(\frac{1+e_0}{1+e_j}\right)\right].\end{aligned}$$

The result follows. $\square$

We now do the critical case, which is almost identical.

Proof of Theorem 3.2.2. By Proposition 3.5.2, the *ordered* split times are independent of $\mathcal{H}$ and have density

$$\begin{aligned}f_k^T(s_1,\ldots,s_{k-1})&=\frac{k!(\beta T+1)^k}{T^{k-1}}\int_0^1(1-y)^{k-1}\prod_{j=0}^{k-1}\frac{1}{(\beta(1-y)(T-s_j)+1)^2}\mathrm{d}y.\\ &=\frac{k!(\beta T+1)^k}{T^{k-1}}\int_0^1\frac{1}{(1+\theta\beta T)^2}\prod_{j=1}^{k-1}\frac{\theta}{(1+\theta\beta(T-s_j))^2}\mathrm{d}\theta.\end{aligned}$$

for any $0 \le s_1 \le \ldots \le s_{k-1} \le 1$, where $s_0 = 0$. Therefore (see Lemma 3.7.3) the *unordered* split times are independent of $\mathcal{H}$ and have density

$$\tilde{f}_k^T(s_1, \ldots, s_{k-1}) = \frac{k(\beta T + 1)^k}{T^{k-1}} \int_0^1 \frac{1}{(1+\theta\beta T)^2} \prod_{j=1}^{k-1} \frac{\theta}{(1+\theta\beta(T - s_j))^2} \mathrm{d}\theta.$$

Integrating over s_j for each $j = 1, \ldots, k-1$, we get

$$\begin{aligned}
&\mathbb{P}(\tilde{\mathcal{S}}_1 \ge s_1, \ldots, \tilde{\mathcal{S}}_{k-1} \ge s_{k-1} | N_T \ge k)\\
&\qquad = k\beta T\Big(1 + \frac{1}{\beta T}\Big)^k \int_0^1 \frac{1}{(1+\theta\beta T)^2} \prod_{j=1}^{k-1} \Big(1 - \frac{1}{1+\theta\beta(T-s_j)}\Big) \mathrm{d}\theta\\
&\qquad = kT\Big(1 + \frac{1}{\beta T}\Big)^k \int_0^1 \frac{1}{(1+\theta T)^2} \prod_{j=1}^{k-1} \Big(1 - \frac{1}{1+\theta(T-s_j)}\Big) \mathrm{d}\theta.
\end{aligned}$$

We can now apply the second part of Lemma 3.7.1, with $e_j = (T - s_j)$ and $s_0 = 0$. This gives

$$\begin{aligned}
&\mathbb{P}(\tilde{\mathcal{S}}_1 \ge s_1, \ldots, \tilde{\mathcal{S}}_{k-1} \ge s_{k-1} | N_T \ge k)\\
&\quad = kT\Big(1 + \frac{1}{\beta T}\Big)^k \left[\frac{1}{1+e_0} \prod_{i=1}^{k-1} \frac{e_i}{e_i - e_0} - \sum_{j=1}^{k-1} \frac{e_j}{(e_j - e_0)^2} \Big(\prod_{\substack{i=1\\ i \ne j}}^{k-1} \frac{e_i}{e_i - e_j} \Big) \log\Big(\frac{1+e_0}{1+e_j}\Big) \right].
\end{aligned}$$

The result now follows from some simple manipulation. $\square$

3.6. The near-critical scaling limit

We now let our offspring distribution depend on T, writing $\mathbb{P}_T$ in place of $\mathbb{P}$. We suppose that $m_T := \mathbb{P}_T[L] = 1 + \mu/T + o(1/T)$ for some $\mu \in \mathbb{R}$, and $\mathbb{P}_T[L(L-1)] = \sigma + o(1)$ for some $\sigma > 0$. We also assume that L^2 is uniformly integrable (that is, for all $\varepsilon > 0$ there exists M such that $\sup_T \mathbb{P}_T[L^2 \mathbb{1}_{\{L \ge M\}}] < \varepsilon$). We define $\mathbb{Q}_T^{k,T}$ just as before, except that it is defined relative to $\mathbb{P}_T^k$ instead of $\mathbb{P}^k$.

In order to prove our results we would like some conditions on the higher moments of L. The next lemma ensures that we may make some further assumptions without loss of generality.

Lemma 3.6.1. Fix $k \ge 1$. Under $\mathbb{P}_T$, there exists a coupling between our Galton-Watson tree with offspring distribution L (and its k chosen particles) and another Galton-Watson tree with offspring distribution $\tilde{L}$ satisfying

- $\mathbb{P}_T[\tilde{L}] = 1 + \mu/T + o(1/T)$;
- $\mathbb{P}_T[\tilde{L}(\tilde{L} - 1)] = \sigma + o(1)$;
- $\mathbb{P}_T[\tilde{L}^{(j)}] = o(T^{j-2})$ for all $j \ge 3$,

such that for each k, with probability tending to 1, the two trees induced by the k chosen particles are equal until time T.

The proof of this lemma is interesting, but not really relevant to the rest of our investigation, so we have included it in the appendix.

In light of Lemma 3.6.1, we further assume without loss of generality that our offspring distribution L satisfies

$$\mathbb{P}_T[L^{(j)}] = o(T^{j-2}) \quad \forall j = 3, \ldots, k. \tag{3.19}$$

3.6.1 Estimating moments and generating functions under $\mathbb{P}$

In Section 3.5.1, we calculated generating functions and moments of the population size under $\mathbb{P}$ precisely for birth-death processes. With more complicated offspring distributions this is no longer possible, but the near-criticality ensures that we can give good approximations.

Lemma 3.6.2. For $k \geq 1$, the kth descending moment $M_k(t) = \mathbb{E}[N_t^{(k)}]$ of any continuous-time Galton-Watson process satisfies

$$M_k'(t) = kr(m-1)M_k(t) + r\sum_{j=2}^{k}\binom{k}{j}\mathbb{E}[L^{(j)}]\,M_{k+1-j}(t).$$

Proof. As before let $F(\theta, t) = \mathbb{E}[\theta^{N_t}]$, and let $u(\theta) = \mathbb{E}[\theta^L] - \theta$. Then F and u satisfy the Kolmogorov forward equation (3.8)

$$\frac{\partial F(\theta,t)}{\partial t} = ru(\theta)\frac{\partial F(\theta,t)}{\partial \theta}. \tag{3.20}$$

Note that

$$M_k(t) = \Big[\frac{\partial^k}{\partial\theta^k}F(\theta,t)\Big]_{\theta=1}, \tag{3.21}$$

so, using the fact that F is smooth,

$$\frac{d}{dt}M_k(t) = \frac{d}{dt}\Big[\frac{\partial^k}{\partial\theta^k}F(\theta,t)\Big]_{\theta=1} = \Big[\frac{\partial}{\partial t}\frac{\partial^k}{\partial\theta^k}F(\theta,t)\Big]_{\theta=1} = \Big[\frac{\partial^k}{\partial\theta^k}\frac{\partial}{\partial t}F(\theta,t)\Big]_{\theta=1}.$$

Applying (3.20),

$$\frac{d}{dt}M_k(t) = \Big[\frac{\partial^k}{\partial\theta^k}\Big(ru(\theta)\frac{\partial F}{\partial\theta}\Big)\Big]_{\theta=1}$$

so using (3.21) again,

$$\frac{d}{dt}M_k(t) = r\sum_{j=0}^{j=k}\binom{k}{j}u^{(j)}(1)\,M_{k+1-j}(t).$$

Finally, $u(1) = 0$, $u'(1) = (m-1)$, and $u^{(j)}(1) = \mathbb{E}[L^{(j)}]$ for $j \geq 2$. $\square$

For real-valued functions f and g, we write $f(x) = o(g(x))$ to mean that $f(x)/g(x) \to 0$ as $x \to \infty$.

Lemma 3.6.3. If $\mu \neq 0$ then the descending moments at scaled times satisfy

$$\lim_{T\to\infty}\frac{\mathbb{P}_T[N_{sT}^{(k)}]}{T^{k-1}} = \Big(\frac{\sigma}{2\mu}\Big)^{k-1}k!e^{r\mu s}(e^{r\mu s}-1)^{k-1}$$

for all $k \geq 1$ and $s \in [0,1]$. If $\mu = 0$ then instead

$$\lim_{T\to\infty}\frac{\mathbb{P}_T[N_{sT}^{(k)}]}{T^{k-1}} = k!\Big(\frac{r\sigma s}{2}\Big)^{k-1}$$

for all $k \geq 1$ and $s \in [0,1]$.

Proof. We proceed by induction. Note that both statements are true for $k = 1$. Letting $M_k(t) = \mathbb{P}_T[N_t^{(k)}]$, by Lemma 3.6.2 we have

$$M_k'(t) = kr(m_T - 1)M_k(t) + r\sum_{j=2}^{k}\binom{k}{j}\mathbb{P}_T[L^{(j)}]\, M_{k+1-j}(t).$$

So letting $\hat{M}_k(s) = M_k(sT)$, we have

$$\begin{aligned}\hat{M}_k'(s) &= T\Big(kr(m_T - 1)\hat{M}_k(s) + r\sum_{j=2}^{k}\binom{k}{j}\mathbb{P}_T[L^{(j)}]\, \hat{M}_{k+1-j}(s)\Big)\\ &= kr\mu\hat{M}_k(s) + Tr\binom{k}{2}\sigma\hat{M}_{k-1}(s) + o(T^{k-1}) \qquad (3.22)\end{aligned}$$

where we used the induction hypothesis to get the last equality.

We now consider the cases $\mu \neq 0$ and $\mu = 0$ separately. In the case $\mu \neq 0$, using the integrating factor $e^{-kr\mu s}$, and applying the induction hypothesis again, we get

$$\frac{\mathrm{d}}{\mathrm{d}s}\big(e^{-kr\mu s}\hat{M}_k(s)\big) = T^{k-1}k!(k-1)r\mu\Big(\frac{\sigma}{2\mu}\Big)^{k-1}e^{-(k-1)r\mu s}(e^{r\mu s} - 1)^{k-2} + e^{-kr\mu s}O(T^{k-2}). \quad (3.23)$$

Noting that

$$(k-1)r\mu e^{-(k-1)r\mu s}(e^{r\mu s} - 1)^{k-2} = \frac{\mathrm{d}}{\mathrm{d}s}\big(e^{-(k-1)r\mu s}(e^{r\mu s} - 1)^{k-1}\big),$$

by integrating (3.23) we obtain

$$e^{-kr\mu s}\hat{M}_k(s) = T^{k-1}k!\Big(\frac{\sigma}{2\mu}\Big)^{k-1}e^{-(k-1)r\mu s}(e^{r\mu s} - 1)^{k-1} + e^{-kr\mu s}O(T^{k-2}).$$

Multiplying through by $e^{kr\mu s}$ gives the result for $\mu \neq 0$.

If $\mu = 0$, then from (3.22) and the induction hypothesis, we have

$$\hat{M}_k'(s) = T^{k-1}k!\Big(\frac{r\sigma}{2}\Big)^{k-1}(k-1)s^{k-2} + o(T^{k-1})$$

and integrating directly gives the result. $\square$

3.6.2 Asymptotics for the generating function

Define

$$F_T(\theta, t) = \mathbb{P}_T[\theta^{N_t}], \qquad u_T(\theta) = \mathbb{P}_T[\theta^L] - \theta,$$

and

$$f_T(\phi, s) = T\big(1 - \mathbb{P}_T[e^{-\frac{\phi}{T}N_{sT}}]\big) = T(1 - F_T(e^{-\phi/T}, sT)).$$

The following result will be important for approximating terms that arise from Campbell's formula.

Lemma 3.6.4. For each $\phi \geq 0$,

$$f_T(\phi, s) \to f(\phi, s)$$

and

$$T^2 u_T(F_T(e^{-\phi/T}, sT)) \to -\mu f(\phi,s) + \frac{\sigma}{2} f(\phi,s)^2$$

as $T \to \infty$, uniformly over $s \in [0,1]$, where

$$f(\phi,s) = \frac{\phi e^{\mu r s}}{1+\frac{\sigma}{2\mu}\phi(e^{\mu r s}-1)} \qquad \text{if } \ \mu \neq 0$$

and

$$f(\phi,s) = \frac{\phi}{1+r\sigma\phi s/2} \qquad \text{if } \ \mu = 0.$$

Proof. First we show that for each ϕ, f_T is bounded in $T > 0$ and $s \in [0,1]$. Note that $x \mapsto 1-e^{-\kappa x}$ is concave and increasing for any $\kappa \geq 0$, so by Jensen's inequality,

$$f_T(\phi,s) = T\big(1-\mathbb{P}[e^{-\frac{\phi}{T}N_{sT}}]\big) \leq T\big(1-e^{-\frac{\phi}{T}\mathbb{P}_T[N_{sT}]}\big) \leq T\big(1-e^{-\frac{\phi}{T}\exp(r\mu+o(1))}\big).$$

Applying the inequality $1-e^{-x} \leq x$, we see that

$$f_T(\phi,s) \leq \phi e^{r\mu+o(1)}.$$

Now, with $F_T(\theta,t) = \mathbb{P}_T[\theta^{N_t}]$, we have

$$\frac{\partial f_T(\phi,s)}{\partial s} = \frac{\partial}{\partial s}\big(T(1-F_T(e^{-\phi/T},sT))\big) = -T^2\frac{\partial F_T(e^{-\phi/T},t)}{\partial t}\Big|_{t=sT}. \tag{3.24}$$

By the Kolmogorov backwards equation (3.9),

$$\frac{\partial}{\partial t}F_T(\theta,t) = ru_T(F_T(\theta,t)) = r\mathbb{P}_T[F_T(\theta,t)^L] - rF_T(\theta,t), \tag{3.25}$$

so

$$\frac{\partial f_T(\phi,s)}{\partial s} = T^2 r\sum_{j=0}^{\infty} p_j^{(T)}\big(F(e^{-\phi/T},sT) - F(e^{-\phi/T},sT)^j\big) = T^2 r\sum_{j=0}^{\infty} p_j^{(T)}\Big(1-\frac{f_T}{T} - \Big(1-\frac{f_T}{T}\Big)^j\Big)$$

where $p_j^{(T)} = \mathbb{P}_T(L=j)$. Expanding $(1-f_T/T)^j$, we get

$$\begin{aligned}\frac{\partial f_T(\phi,s)}{\partial s} &= T^2 r\sum_{j=0}^{\infty} p_j^{(T)}\Big((j-1)\frac{f_T}{T} - \frac{j(j-1)f_T^2}{2T^2} - \sum_{i=3}^{j}\binom{j}{i}\Big(-\frac{f_T}{T}\Big)^i\Big)\\ &= r\mu f_T - \frac{r\sigma}{2}f_T^2 + o(1) - T^2 r\sum_{j=0}^{\infty} p_j^{(T)}\sum_{i=3}^{j}\binom{j}{i}\Big(-\frac{f_T}{T}\Big)^i.\end{aligned}$$

Swapping the order of summation, this becomes

$$\begin{aligned}\frac{\partial f_T(\phi,s)}{\partial s} &= r\mu f_T - \frac{r\sigma}{2}f_T^2 + o(1) - T^2 r\sum_{i=2}^{\infty}\frac{1}{i!}\Big(-\frac{f_T}{T}\Big)^i\sum_{j=i}^{\infty} p_j^{(T)} j(j-1)\dots(j-i+1)\\ &= r\mu f_T - \frac{r\sigma}{2}f_T^2 + o(1) - T^2 r\sum_{i=2}^{\infty}\frac{1}{i!}\Big(-\frac{f_T}{T}\Big)^i\mathbb{P}_T[L^{(i)}]\\ &= r\mu f_T - \frac{r\sigma}{2}f_T^2 + o(1)\end{aligned} \tag{3.26}$$

since f_T is bounded and $\mathbb{P}_T[L^{(i)}] = o(T^{i-2})$ for each $i \geq 3$ (see (3.19)). Note in particular that the $o(1)$ term is uniform in s.

Note that f is the solution to

$$\frac{\partial f}{\partial s} = r\mu f - \frac{r\sigma}{2} f^2$$

with $f(\phi, 0) = \phi$. Setting $h_T(\phi, s) = f_T(\phi, s) - f(\phi, s)$ we have

$$\frac{\partial h_T}{\partial s} = r\mu(f_T - f) - \frac{r\sigma}{2}(f_T^2 - f^2) + o(1)$$

where the $o(1)$ term is uniform in s. Integrating over s with ϕ fixed,

$$h_T(\phi, s) = h_T(\phi, 0) + r\mu \int_0^s h_T(\phi, s')\mathrm{d}s' - \frac{r\sigma}{2} \int_0^s h_T(\phi, s')(f_T(\phi, s') + f(\phi, s'))\mathrm{d}s + o(1).$$

For fixed ϕ, both f_T and f are bounded in s and T, say by M_ϕ. Also $|h_T(\phi, 0)| = T(1 - e^{-\phi/T}) - \phi = o(1)$. Thus

$$|h_T(\phi, s)| \leq r \int_0^s |h_T(\phi, s')|(\mu + \sigma M_\phi/2)\mathrm{d}s' + o(1),$$

where again the $o(1)$ term is uniform in s. Gronwall's inequality then tells us that $|h_T(\phi, s)| \to 0$ uniformly in s. This proves the first part of the lemma.

The second part of the lemma is now implicit in our calculations above: by (3.25) and then (3.24),

$$u_T(F_T(e^{-\phi/T}, sT)) = \frac{1}{r}\frac{\partial}{\partial t} F_T(e^{-\phi/T}, t)|_{t=sT} = -\frac{1}{rT^2}\frac{\partial f_T(\phi, s)}{\partial s}.$$

Applying (3.26) tells us that

$$T^2 u_T(F_T(e^{-\phi/T}, sT)) = -\mu f_T + \frac{\sigma}{2} f_T^2 + o(1),$$

and by the first part of the lemma we get

$$T^2 u_T(F_T(e^{-\phi/T}, sT)) \to -\mu f + \frac{\sigma}{2} f^2. \qquad \square$$

Lemma 3.6.5. For any $s \in (0, 1]$, as $T \to \infty$,

$$T\mathbb{P}_T(N_{sT} > 0) \to \frac{2\mu e^{\mu r s}}{\sigma(e^{\mu r s} - 1)} \qquad \text{if } \mu \neq 0$$

and

$$T\mathbb{P}_T(N_{sT} > 0) \to \frac{2}{r\sigma s} \qquad \text{if } \mu = 0.$$

Proof. Note that $\mathbb{P}_T(N_t = 0) = F_T(0, t)$, and so satisfies the Kolmogorov backwards equation (3.9). Thus the proof of Lemma 3.6.4 works exactly the same for

$$T\mathbb{P}_T(N_{sT} > 0) = T(1 - \mathbb{P}_T(N_{sT} = 0)) = T(1 - F_T(0, sT)),$$

except for showing that $T\mathbb{P}_T(N_{sT} > 0)$ is bounded—we can no longer apply Jensen's inequality.

Instead, we note that in the critical case $m_T = 1$ the boundedness is well known (see for example [5, Chapter III, Section 7, Lemma 2]). When $m_T \neq 1$, let $\bar{p}_0^{(T)} = p_0^{(T)}$ and for $j \geq 1$,

$$\bar{p}_j^{(T)} = p_j^{(T)} + (1 - m_T)2^{-j}/j.$$

This gives us a new offspring distribution $\bar{L}$ that is critical (and has finite variance). We can then easily construct a coupling between N_t and $\bar{N}_t$, where $\bar{N}_t$ is the number of particles in a branching process with offspring distribution $\bar{L}$, such that

- if $m_T < 1$, then $N_t \le \bar{N}_t$ for all $t \ge 0$;
- if $m_T > 1$, then $N_t \ge \bar{N}_t$ for all $t \ge 0$.

In the case $m_T < 1$, we have $T\mathbb{P}(N_{sT} > 0) \le T\mathbb{P}(\bar{N}_{sT} > 0)$, which is bounded. In the case $m_T > 1$, we have

$$\mathbb{P}_T(N_{sT} > 0) = \mathbb{Q}_T^{1,sT}\Big[\frac{\mathbb{P}_T[N_{sT}]}{N_{sT}}\Big] = e^{r(m_T-1)sT}\mathbb{Q}_T^{1,sT}\Big[\frac{1}{N_{sT}}\Big]$$

and similarly for $\bar{N}_{sT}$ with its equivalent measure $\bar{\mathbb{Q}}_T^{1,sT}$. Since $T\mathbb{P}(\bar{N}_{sT} > 0)$ is bounded, we get that $T\bar{\mathbb{Q}}_T^{1,sT}[1/\bar{N}_{sT}]$ is bounded, but

$$\mathbb{Q}_T^{1,sT}\Big[\frac{1}{N_{sT}}\Big] \le \bar{\mathbb{Q}}_T^{1,sT}\Big[\frac{1}{\bar{N}_{sT}}\Big],$$

so $T\mathbb{Q}_T^{1,sT}[1/N_{sT}]$ is bounded and therefore $T\mathbb{P}_T(N_{sT} > 0)$ is also bounded. This completes the proof. □

3.6.3 Spine split times under $\mathbb{Q}_T^{k,T}$

We now want to feed our calculations for moments and generating functions under $\mathbb{P}$ into understanding the spine split times under $\mathbb{Q}$, as in Lemma 3.5.1. Unfortunately the spine split times in non-binary cases do not have a joint density with respect to Lebesgue measure: for any $j = 2, \dots, k-1$, there is a positive probability that $\psi_j = \psi_{j-1}$. However we show that this probability tends to zero as $T \to \infty$, and therefore will not have an effect on our final answer.

Recall that n_t is the number of distinct spine particles at time t, and ρ_t^i is the number of marks carried by spine i at time t.

Lemma 3.6.6. For any $i = 1, \dots, k-1$ and $t \in (0,1)$,

$$\mathbb{Q}_T^{k,T}\Big(n_{\psi_1} = 2,\ \rho^1_{\psi_1} = i \,\Big|\, \frac{\psi_1}{T} = t\Big) \to \frac{1}{k-1}.$$

This tells us two things: that with probability tending to 1 we have exactly 2 spines at the first spine split time; and that the number of marks following each of those spines is uniformly distributed on $1, \dots, k-1$.

Proof. We work in the case $\mu \ne 0$; the case $\mu = 0$ proceeds almost identically. From the definition of $\mathbb{Q}$,

$$\mathbb{Q}_T^{k,T}(n_{tT} = 2,\ \rho^1_{tT} = i \,|\, \tau_\emptyset = tT,\ n_{tT} \ge 2) = \frac{\mathbb{P}_T[g_{k,T}\mathbb{1}_{\{n_{tT}=2,\, \rho^1_{tT}=i\}}|\tau_\emptyset = tT]}{\mathbb{P}_T[g_{k,T}\mathbb{1}_{\{n_{tT}\ge 2\}}|\tau_\emptyset = tT]}.$$

Let $P_T(j; b; a_1, \dots, a_b)$ be the probability that at time $\tau_\emptyset$, j children are born, b of which are spines, carrying $a_1, \dots, a_b$ marks. Then

$$\mathbb{P}_T[g_{k,T}\mathbb{1}_{\{n_{tT}=b,\, \rho^1_{tT}=a_1\}} \,|\, \tau_\emptyset = tT] = \sum_{j=b}^{\infty} \sum_{a_2,\dots,a_b} P_T(j; b; a_1, \dots, a_b) j^k \prod_{i=1}^{b} \mathbb{P}_T[g_{a_i, T(1-t)}]$$

where the sum over $a_2, \dots, a_b$ runs over $1, \dots, k$ such that $a_1 + \dots + a_b = k$. Now

$$P_T(j; b; a_1, \dots, a_b) = p_j^{(T)} \binom{j}{b} \frac{k!}{a_1! \dots a_b!} \frac{1}{j^k}$$

and from Lemma 3.6.3, in the case $\mu \neq 0$,

$$\mathbb{P}_T[N^{(a_i)}_{T(1-t)}] = T^{a_i-1}\Big(\frac{\sigma}{2\mu}\Big)^{a_i-1} a_i! e^{r\mu(1-t)}(e^{r\mu(1-t)}-1)^{a_i-1} + o(T^{a_i-1}).$$

This gives us

$$\begin{aligned}\mathbb{P}_T[g_{k,T}\mathbb{1}_{\{n_{tT}=b,\,\rho^1_{tT}=a_1\}}\,|\,\tau_\emptyset = tT]&\\ &= \sum_{j=b}^{\infty}\sum_{a_2,\ldots,a_b} p_j^{(T)}\binom{j}{b} k! T^{k-b}\Big(\frac{\sigma}{2\mu}\Big)^{k-b} e^{br\mu(1-t)}(e^{r\mu(1-t)}-1)^{k-b}(1+o(1)).\end{aligned}$$

If $b=2$, then fixing $a_1 = i$ also fixes a_2 since $a_2 = k - a_1$, so the second sum disappears and we are left with

$$\begin{aligned}\mathbb{P}_T[g_{k,T}\mathbb{1}_{\{n_{tT}=2,\,\rho^1_{tT}=i\}}|\tau_\emptyset = tT] &= \sum_{j=2}^{\infty} p_j^{(T)}\binom{j}{2} k! T^{k-2}\Big(\frac{\sigma}{2\mu}\Big)^{k-2} e^{2r\mu(1-t)}(e^{r\mu(1-t)}-1)^{k-2}(1+o(1))\\ &= \frac{\sigma}{2} k! T^{k-2}\Big(\frac{\sigma}{2\mu}\Big)^{k-2} e^{2r\mu(1-t)}(e^{r\mu(1-t)}-1)^{k-2}(1+o(1)).\end{aligned} \tag{3.27}$$

Notice in particular that this does not depend on the value of i.

Next we bound the probability that there are at least three distinct spines at time ψ_1 by taking a sum over a_1 and then over $b \geq 3$. For each b, there are certainly at most k^b possible values of $a_1, \ldots, a_b$ that sum to k. Thus we get

$$\mathbb{P}_T[g_{k,T}\mathbb{1}_{\{n_{tT}\geq 3\}}\,|\,\tau_\emptyset = tT] \leq \sum_{b=3}^{\infty}\mathbb{P}_T[L^{(b)}]\frac{k!}{b!}k^b T^{k-b}\Big(\frac{\sigma}{2\mu}\Big)^{k-b} e^{br\mu(1-t)}(e^{r\mu(1-t)}-1)^{k-b}(1+o(1)).$$

Recall that we have assumed (3.19) that $\mathbb{P}_T[L^{(b)}] = o(T^{b-2})$ for each $b \geq 3$, so

$$\mathbb{P}_T[g_{k,T}\mathbb{1}_{\{n_{tT}\geq 3\}}\,|\,\tau_\emptyset = tT] = o(T^{k-2}). \tag{3.28}$$

Dividing (3.28) by (3.27), we see that the probability that there are at least 3 distinct spines at time ψ_1 tends to zero as $T \to \infty$; or equivalently, that the probability that there are exactly 2 distinct spines tends to 1. Then since the right-hand side of (3.27) does not depend on i, the distribution of ρ_{ψ_1} must be asymptotically uniform. □

Combined with the symmetry lemma, the previous result tells us that with high probability the spine split times are distinct. We want to use this to show that away from 0, the rescaled split times $\psi_1/T, \ldots, \psi_{k-1}/T$ have an asymptotic density. First we need a preparatory lemma, which will be helpful in describing the topology of our limiting tree as well as calculating the asymptotic density of the split times.

Lemma 3.6.7. For any $s \in (0,1]$ and $t \in (0,s)$,

$$\mathbb{Q}_T^{k,sT}\Big(\frac{\psi_1}{T} > t\Big) \to \Big(\frac{e^{r\mu(s-t)}-1}{e^{r\mu s}-1}\Big)^{k-1}$$

and

$$-\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{Q}_T^{k,sT}\Big(\frac{\psi_1}{T} > t\Big) \to (k-1)r\mu\frac{(e^{r\mu(s-t)}-1)^{k-2}}{(e^{r\mu s}-1)^{k-1}}e^{r\mu(s-t)}$$

as $T \to \infty$.

Proof. The first part of the proof follows easily by combining Lemmas 3.4.7 and 3.6.3. The second part is a more involved calculation. As in Lemma 3.6.2, we write $M_k(t) = \mathbb{P}_T[N_t^{(k)}]$. By Lemma 3.4.7,

$$\mathbb{Q}_T^{k,sT}(\psi_1 > tT) = \frac{\mathbb{P}[N_{T(s-t)}^{(k)}]}{\mathbb{P}[N_{sT}^{(k)}]} e^{(m_T-1)rtT} = \frac{M_k(T(s-t))}{M_k(sT)} e^{(m_T-1)rtT},$$

so

$$\begin{aligned}
-\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{Q}_T^{k,sT}(\psi_1 > tT) &= T\frac{M_k'(T(s-t))}{M_k(sT)}e^{(m_T-1)rtT} - T(m_T-1)r\frac{M_k(T(s-t))}{M_k(sT)}e^{(m_T-1)rtT} \\
&= \frac{T}{M_k(sT)}e^{(m_T-1)rtT}\big(M_k'(T(s-t)) - (m_T-1)rM_k(T(s-t))\big).
\end{aligned}$$

Applying Lemma 3.6.2, this equals

$$\frac{T}{M_k(sT)}e^{(m_T-1)rtT}\Bigg((k-1)r(m_T-1)M_k(T(s-t)) + r\sum_{j=2}^{k}\binom{k}{j}\mathbb{P}_T[L^{(j)}]M_{k+1-j}(T(s-t))\Bigg).$$

We now use Lemma 3.6.3. Since $\mathbb{P}_T[L^{(j)}] = o(T^{j-2})$ for all $j \geq 3$ (see (3.19)), the terms with $j \geq 3$ in the sum above do not contribute in the limit. We obtain

$$\begin{aligned}
\frac{Te^{r\mu t}}{(\frac{\sigma}{2\mu})^{k-1}k!e^{r\mu s}(e^{r\mu s}-1)^{k-1}T^{k-1}}\Bigg[&(k-1)r\mu\Big(\frac{\sigma}{2\mu}\Big)^{k-1}k!e^{r\mu(s-t)}(e^{r\mu(s-t)}-1)^{k-1}T^{k-2} \\
&+ r\frac{k(k-1)}{2}\sigma\Big(\frac{\sigma}{2\mu}\Big)^{k-2}(k-1)!e^{r\mu(s-t)}(e^{r\mu(s-t)}-1)^{k-2}T^{k-2} + o(T^{k-2})\Bigg].
\end{aligned}$$

Simplifying, this equals

$$\frac{1}{(e^{r\mu s}-1)^{k-1}}\Big[(k-1)r\mu(e^{r\mu(s-t)}-1)^{k-1} + (k-1)r\mu(e^{r\mu(s-t)}-1)^{k-2} + o(1)\Big],$$

so simplifying again we get

$$-\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{Q}_T^{k,sT}(\psi_1 > tT) \to (k-1)r\mu\frac{(e^{r\mu(s-t)}-1)^{k-2}}{(e^{r\mu s}-1)^{k-1}}e^{r\mu(s-t)}. \qquad \square$$

Recall that $\mathcal{H}'$ is the σ-algebra containing topological information about which marks are following which spines, without information about the spine split times.

Proposition 3.6.8. The spine split times $\psi_1, \dots, \psi_{k-1}$ are asymptotically independent of $\mathcal{H}'$ under $\mathbb{Q}_T^{k,T}$, and for any $0 < s_1 < t_1 \leq s_2 < t_2 \leq \ldots \leq s_{k-1} < t_{k-1} < 1$,

$$\lim_{T\to\infty}\mathbb{Q}_T^{k,T}\Big(\frac{\psi_1}{T} \in (s_1, t_1], \dots, \frac{\psi_{k-1}}{T} \in (s_{k-1}, t_{k-1}]\Big) = \int_{s_1}^{t_1}\cdots\int_{s_{k-1}}^{t_{k-1}} f_k(s_1', \dots, s_{k-1}')\,\mathrm{d}s_{k-1}'\ldots\mathrm{d}s_1',$$

where

$$f_k(s_1, \dots, s_{k-1}) = (k-1)!\Big(\frac{r\mu}{e^{r\mu}-1}\Big)^{k-1}\prod_{i=1}^{k-1}e^{r\mu(1-s_i)} \qquad \text{if } \mu \neq 0$$

and

$$f_k(s_1, \dots, s_{k-1}) = (k-1)! \qquad \text{if } \mu = 0.$$

Proof. This is a generalization of the proof of Lemma 3.5.1, and the reader may wish to compare the two. The main difference is that now there is a chance that spine splitting events result in more than one new spine particle (since branching events need not be binary), and therefore we need to take care over ensuring that the split times $\psi_1, \dots, \psi_{k-1}$ are distinct.

With this in mind, let Υ_j be the event that the first j spine split times are distinct,

$$\Upsilon_j = \{\psi_i \neq \psi_{i-1} \ \forall i = 2, \dots, j\}.$$

We work by induction; fix $j \leq k-1$, $T > 0$, $0 < s_1 < \dots < s_{j-1} < 1$. Then for $s \geq s_{j-1}$,

$$\begin{aligned}
&\mathbb{Q}\Big(\frac{\psi_j}{T} > s \,\Big|\, \frac{\psi_{j-1}}{T} = s_{j-1}, \dots, \frac{\psi_1}{T} = s_1\Big) \\
&= \mathbb{Q}\Big(\Upsilon_j,\, \frac{\psi_j}{T} > s \,\Big|\, \frac{\psi_{j-1}}{T} = s_{j-1}, \dots, \frac{\psi_1}{T} = s_1\Big) \\
&= \mathbb{Q}\Big(\frac{\psi_j}{T} > s \,\Big|\, \Upsilon_j,\, \frac{\psi_{j-1}}{T} = s_{j-1}, \dots, \frac{\psi_1}{T} = s_1\Big)\mathbb{Q}\Big(\Upsilon_j \,\Big|\, \frac{\psi_{j-1}}{T} = s_{j-1}, \dots, \frac{\psi_1}{T} = s_1\Big).
\end{aligned}$$

By Lemma 3.6.6 and the symmetry lemma,

$$\mathbb{Q}\Big(\Upsilon_j \,\Big|\, \frac{\psi_{j-1}}{T} = s_{j-1}, \dots, \frac{\psi_1}{T} = s_1\Big) \to 1$$

for all $0 < s_1 < \dots < s_{j-1} < 1$. We also set

$$D(s) = -\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{Q}\Big(\frac{\psi_j}{T} > s \,\Big|\, \Upsilon_j,\, \frac{\psi_{j-1}}{T} = s_{j-1}, \dots, \frac{\psi_1}{T} = s_1\Big)$$

and claim that

$$D(s) = (k-j)r\mu e^{r\mu(1-s_j)}\frac{(e^{r\mu(1-s_j)} - 1)^{k-j-1}}{(e^{r\mu(1-s_{j-1})} - 1)^{k-j}} + o(1).$$

If this claim holds, then applying induction and taking a product over j gives the result. In particular, since this does not depend on the number of marks following each spine, the split times are asymptotically independent of $\mathcal{H}'$.

To prove the claim, fix $a_1, \dots, a_j$ such that $a_i \in \{1, \dots, k\}$ for each i and $a_1 + \dots + a_j = k$. Let A_j be the event that after time ψ_{j-1}, we have j distinct spine particles carrying $a_1, \dots, a_j$ marks. Then by the symmetry lemma (letting $s_0 = 0$),

$$\mathbb{Q}_T^{k,T}\Big(\frac{\psi_j}{T} > s_j \,\Big|\, \Upsilon_j,\, A_j,\, \frac{\psi_{j-1}}{T} = s_{j-1}\Big) = \prod_{i=1}^{j} \mathbb{Q}_T^{a_i, T(1-s_{j-1})}(\psi_1/T > s_j - s_{j-1}).$$

Thus, differentiating, we have

$$D(s) = -\sum_{a_1,\dots,a_j} P_{a_1,\dots,a_j} \sum_{l=1}^{j} \Big(\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{Q}_T^{a_l, T(1-s_{j-1})}(\psi_1/T > s - s_{j-1})\Big)\prod_{i\neq l}\mathbb{Q}_T^{a_i, T(1-s_{j-1})}(\psi_1/T > s - s_{j-1})$$

where $P_{a_1,\dots,a_j}$ is the probability that A_j occurs. Applying Lemma 3.6.7 then establishes the claim and completes the proof. $\square$

We recall now the partition $Z_0, Z_1, \dots$ which contained the information about the marks following each of the distinct spine particles, without the information about the split times.

Lemma 3.6.9. The partition $Z_0, Z_1, \dots$ has the following distribution under $\mathbb{Q}_T^{k,T}$:

- If Z_i consists of $i+1$ blocks of sizes $a_1, \dots, a_{i+1}$, then the jth block will split next with probability $\frac{a_j-1}{k-i-1}(1+o(1))$ for each $j = 1, \dots, i+1$.
- When a block of size a splits, it splits into two new blocks with probability $1+o(1)$, and the probability that these blocks have sizes l and $a-l$ is $\frac{1}{a-1}(1+o(1))$ for each $l = 1, \dots, a-1$.

Proof. Suppose that we are given $\psi_i = sT$. For the first part, by the symmetry lemma, the probability that the jth block splits next is

$$\int_0^{T(1-s)} \mathbb{Q}_T^{a_j,T(1-s)}\Big(\frac{\psi_1}{T} \in \mathrm{d}t\Big) \prod_{l\neq j} \mathbb{Q}_T^{a_l,T(1-s)}\Big(\frac{\psi_1}{T} > t\Big)$$
$$= \int_0^{T(1-s)} \Big(-\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{Q}_T^{a_j,T(1-s)}\Big(\frac{\psi_1}{T} > t\Big)\Big) \prod_{l\neq j} \mathbb{Q}_T^{a_l,T(1-s)}\Big(\frac{\psi_1}{T} > t\Big)\,\mathrm{d}t.$$

By Lemma 3.6.7, this converges as $T \to \infty$ to

$$(a_j - 1)r\mu \int_0^{T(1-s)} e^{r\mu(1-s-t)} \frac{e^{(r\mu(1-s-t)} - 1)^{k-i}}{e^{(r\mu(1-s)} - 1)^{k-i-1}}\mathrm{d}t.$$

Since the integrand does not depend on a_j, and we know the sum of the above quantity over $j = 1, \dots, i+1$ must converge to 1 (since one of the blocks must split first), we get

$$r\mu \int_0^{T(1-s)} e^{r\mu(1-s-t)} \frac{e^{(r\mu(1-s-t)} - 1)^{k-b-1}}{e^{(r\mu(1-s)} - 1)^{k-b}}\mathrm{d}t \to \frac{1}{k-i-1}$$

and therefore the probability that the jth block splits next converges to $\frac{a_j-1}{k-i-1}$ as claimed.

The second part follows immediately from Lemma 3.6.6. □

3.6.4 Asymptotics for N_T under $\mathbb{Q}_T^{k,T}$

We now apply our asymptotics for $u_T(F_T(e^{-z}, sT))$ to approximate the distribution of N_T when the split times are known.

Lemma 3.6.10. For any $\phi \geq 0$ and $0 \leq s_1 \leq \ldots \leq s_{k-1} \leq 1$,

$$\mathbb{Q}_T^{k,T}\Big[e^{-\phi \tilde{N}_T/T} \,\Big|\, \mathcal{G}_T^k, \frac{\psi_1}{T} = s_1, \dots, \frac{\psi_{k-1}}{T} = s_k\Big] \to \begin{cases} \prod_{i=0}^{k-1} \Big(1 + \frac{\sigma}{2\mu}\phi(e^{r\mu(1-s_i)} - 1)\Big)^{-2} & \text{if } \mu \neq 0 \\ \prod_{i=0}^{k-1} \Big(1 + \frac{r\sigma}{2}\phi(1-s_i)\Big)^{-2} & \text{if } \mu = 0 \end{cases}$$

almost surely as $T \to \infty$.

Proof. From Proposition 3.4.10 we know that

$$\mathbb{Q}_T^{k,T}\Big[e^{-\phi \tilde{N}_T/T}\Big|\mathcal{G}_T^k, \frac{\psi_1}{T} = s_1, \dots, \frac{\psi_{k-1}}{T} = s_k\Big] = \prod_{i=0}^{k-1} \Big(e^{-r(m_T-1)T(1-s_i)} \frac{u_T(F_T(e^{-\phi/T}, T(1-s_i)))}{u_T(e^{-\phi/T})}\Big).$$

Of course $(m_T - 1)T \to \mu$, and Lemma 3.6.4 tells us that

$$T^2 u_T(F_T(e^{-\phi/T}, T(1-s_i))) \to -\mu f(\phi, 1-s_i) + \frac{\sigma}{2} f(\phi, 1-s_i)^2$$

where

$$f(\phi,s)=\frac{\phi e^{\mu rs}}{1+\frac{\sigma}{2\mu}\phi(e^{\mu rs}-1)}\quad\text{if } \mu\neq 0\qquad\text{or}\qquad f(\phi,s)=\frac{\phi}{1+\frac{r\sigma}{2}\phi s}\quad\text{if } \mu=0.$$

Noting that $u_T(e^{-\phi/T})=u_T(F_T(e^{-\phi/T},0))$, we see that

$$e^{-r(m_T-1)T(1-s_i)}\frac{u_T(F_T(e^{-\phi/T},T(1-s_i)))}{u_T(e^{-\phi/T})}\longrightarrow e^{-r\mu(1-s_i)}\frac{-\mu f(\phi,1-s_i)+\frac{\sigma}{2}f(\phi,1-s_i)^2}{-\mu f(\phi,0)+\frac{\sigma}{2}f(\phi,0)^2}.$$

Now, in the case $\mu\neq 0$, we simply write out

$$\begin{aligned}-\mu f(\phi,1-s_i)+\frac{\sigma}{2}f(\phi,1-s_i)^2&=\frac{-\mu\phi e^{r\mu(1-s_i)}(1+\frac{\sigma}{2\mu}\phi(e^{\mu r(1-s_i)}-1))+\frac{\sigma}{2}\phi^2e^{2r\mu(1-s_i)}}{(1+\frac{\sigma}{2\mu}\phi(e^{\mu r(1-s_i)}-1))^2}\\&=\frac{-\mu\phi e^{r\mu(1-s_i)}+\frac{\sigma}{2}\phi^2e^{r\mu(1-s_i)}}{(1+\frac{\sigma}{2\mu}\phi(e^{\mu r(1-s_i)}-1))^2},\end{aligned}$$

so since $-\mu f(\phi,0)+\frac{\sigma}{2}f(\phi,0)^2=-\mu\phi+\sigma\phi^2/2$, we have

$$e^{-r\mu(1-s_i)}\frac{-\mu f(\phi,1-s_i)+\frac{\sigma}{2}f(\phi,1-s_i)^2}{-\mu f(\phi,0)+\frac{\sigma}{2}f(\phi,0)^2}=\Big(1+\frac{\sigma}{2\mu}\phi(e^{\mu r(1-s_i)}-1)\Big)^{-2}.$$

The result in the case $\mu=0$ is very similar. □

Lemma 3.6.11. For any $\phi\geq 0$,

$$\mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k)/T}|\mathcal{G}_T^k]=\mathbb{Q}_T^{k,T}[e^{-\phi\tilde{N}_T/T}|\mathcal{G}_T^k](1+o(1))$$

$\mathbb{Q}_T^{k,T}$-almost surely.

Proof. Recall that $\tilde{N}_T$ is the number of ordinary particles alive at time T, and there are ($\mathbb{Q}$-almost surely) k spines at time T. All other particles are residue particles. Given $\mathcal{G}_T^k$, the number of residue particles is independent of the number of ordinary particles; therefore it suffices to show that

$$\mathbb{Q}[e^{-\phi(N_T-k-\tilde{N}_T)/T}|\mathcal{G}_T^k]\to 1.$$

By Lemma 3.4.9 (and the symmetry lemma) the number of residue particles born at any spine splitting event is stochastically dominated by $\hat{L}_k$. Since non-spine particles behave exactly as under $\mathbb{P}_T$, the number of descendants at time T of any one particle born at time ψ_i is $\mathbb{P}_T[e^{-zN_{T-s}}]|_{s=\psi_i}$. Therefore

$$\mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k-\tilde{N}_T)/T}|\mathcal{G}_T^k]\geq\prod_{i=1}^{k-1}P\Big[\mathbb{P}_T[e^{-\phi N_{T-s}/T}]^{\hat{L}_k}\Big]\Big|_{s=\psi_i}.$$

By Jensen's inequality, for any $t\in[0,T]$,

$$\mathbb{P}_T[e^{-\phi N_t/T}]\geq\exp(-\phi\mathbb{P}_T[N_t]/T)\geq\exp(-\phi e^{r(m_T-1)T}/T),$$

and thus

$$\mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k-\tilde{N}_T)/T}|\mathcal{G}_T^k]\geq P[\exp(-\phi e^{r(m_T-1)T}\hat{L}_k/T)]^{k-1}.$$

The right-hand side converges to 1 as $T\to\infty$, and of course $\mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k-\tilde{N}_T)/T}|\mathcal{G}_T^k]\leq 1$, so we are done. □

Recall that Υ_{k-1} is the event that all the split times are distinct, and $\mathcal{H}'$ is the σ-algebra that contains topological information about which marks follow which spines without information about the spine split times. Let $(\tilde{\psi}_1, \dots, \tilde{\psi}_{k-1})$ be a uniform random permutation of $(\psi_1, \dots, \psi_{k-1})$. We combine several of our results to prove the following.

Lemma 3.6.12. Fix $s_1, \dots, s_{k-1} \in (0,1)$. Let

$$f(\xi_T) = \mathbb{1}_{\{\tilde{\psi}_1/T > s_1, \dots, \tilde{\psi}_{k-1}/T > s_{k-1}, \Upsilon_{k-1}\} \cap H}$$

where $H \in \mathcal{H}'$. There exists a constant h such that $\mathbb{Q}_T^{k,T}(H) \to h$ as $T \to \infty$. For any $\phi \geq 0$, if $\mu \neq 0$ then

$$\lim_{T\to\infty} \mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k)/T} f(\xi_T)] = \Big(\frac{1}{e^{r\mu}-1}\Big)^{k-1} \frac{h}{(1+\frac{\sigma}{2\mu}\phi(e^{r\mu}-1))^2} \prod_{i=1}^{k-1} \frac{e^{r\mu(1-s_i)}-1}{1+\frac{\sigma}{2\mu}\phi(e^{r\mu(1-s_i)}-1)}$$

and if $\mu = 0$ then

$$\lim_{T\to\infty} \mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k)/T} f(\xi_T)] = \frac{h}{(1+r\sigma\phi/2)^2} \prod_{i=1}^{k-1} \frac{1-s_i}{1+r\sigma\phi(1-s_i)/2}.$$

Proof. The fact that $\mathbb{Q}_T^{k,T}(H)$ converges follows from Lemma 3.6.9. Now, by Proposition 3.6.8 and Lemma 3.7.3, in the case $\mu \neq 0$,

$$\begin{aligned}
&\mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k)/T} f(\xi_T)] \\
&\quad = (1+o(1)) \int_{s_1}^1 \cdots \int_{s_{k-1}}^1 \Big(\frac{r\mu}{e^{r\mu}-1}\Big)^{k-1} \Big(\prod_{i=1}^{k-1} e^{r\mu(1-s_i')}\Big) \\
&\qquad\qquad \cdot \mathbb{Q}_T^{k,T}\Big[\mathbb{1}_H \mathbb{Q}_T^{k,T}\Big[e^{-\phi(N_T-k)/T} \,\Big|\, \mathcal{G}_T^k, \frac{\tilde{\psi}_1}{T} = s_1', \dots, \frac{\tilde{\psi}_1}{T} = s_{k-1}'\Big]\Big].
\end{aligned}$$

By Lemma 3.6.11, we may replace $N_T - k$ with $\tilde{N}_T$; and then by Lemma 3.6.10, the above equals

$$\begin{aligned}
&(1+o(1)) \int_{s_1}^1 \cdots \int_{s_{k-1}}^1 \Big(\frac{r\mu}{e^{r\mu}-1}\Big)^{k-1} \Big(\prod_{i=1}^{k-1} e^{r\mu(1-s_i')}\Big) \\
&\qquad\qquad \cdot \mathbb{Q}_T^{k,T}(H) \prod_{j=0}^{k-1} \Big(1 + \frac{\sigma}{2\mu}\phi(e^{r\mu(1-s_j')}-1)\Big)^{-2} \mathrm{d}s_{k-1}' \dots \mathrm{d}s_1'
\end{aligned}$$

almost surely. After some small rearrangements this becomes

$$(1+o(1))\Big(\frac{r\mu}{e^{r\mu}-1}\Big)^{k-1} \frac{h}{(1+\frac{\sigma}{2\mu}\phi(e^{r\mu}-1))^2} \prod_{i=1}^{k-1} \int_{s_i}^1 \frac{e^{r\mu(1-s_i')}}{(1+\frac{\sigma}{2\mu}\phi(e^{r\mu(1-s_i')}-1))^2} \mathrm{d}s_i',$$

and then applying the second part of Lemma 3.7.2 gives the result. The case $\mu = 0$ is similar. □

3.6.5 The final steps in the proof of Theorem 3.2.3

Proof of Theorem 3.2.3. By Proposition 3.4.2, for any measurable f,

$$\mathbb{P}_T\Big[\frac{1}{N_T^{(k)}} \sum_{u \in \mathcal{N}_T^{(k)}} f(u) \,\Big|\, N_T \geq k\Big] = \frac{\mathbb{P}_T[N_T^{(k)}]}{\mathbb{P}_T(N_T \geq k)(k-1)!} \int_0^\infty (e^z-1)^{k-1} \mathbb{Q}_T^{k,T}\Big[e^{-zN_T} f(\xi_T)\Big] \mathrm{d}z.$$

Substituting $z = \phi/T$ and rearranging, we get

$$\frac{1}{(k-1)!}\frac{\mathbb{P}_T[N_T^{(k)}]}{T^{k-1}}\frac{1}{T\mathbb{P}_T(N_T \geq k)}\int_0^\infty (T(1-e^{-\phi/T}))^{k-1}\mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k)/T}f(\xi_T)]\,\mathrm{d}\phi.$$

By Lemma 3.6.3,

$$\frac{\mathbb{P}_T[N_T^{(k)}]}{T^{k-1}} \to \Big(\frac{\sigma}{2\mu}\Big)^{k-1}k!e^{r\mu}(e^{r\mu}-1)^{k-1} \quad \text{if } \mu \neq 0 \quad \text{ and } \quad \frac{\mathbb{P}_T[N_T^{(k)}]}{T^{k-1}} \to \Big(\frac{r\sigma}{2}\Big)^{k-1}k! \quad \text{if } \mu = 0,$$

and by Lemma 3.6.5,

$$T\mathbb{P}_T(N_T \geq k) \to \frac{2\mu e^{r\mu}}{\sigma(e^{r\mu}-1)} \quad \text{if } \mu \neq 0 \quad \text{ and } \quad T\mathbb{P}_T(N_T \geq k) \to \frac{2}{r\sigma} \quad \text{if } \mu \neq 0.$$

Therefore

$$\frac{1}{(k-1)!}\frac{\mathbb{P}_T[N_T^{(k)}]}{T^{k-1}}\frac{1}{T\mathbb{P}_T(N_T \geq k)} \to k\Big(\frac{\sigma}{2\mu}\Big)^k(e^{r\mu}-1)^k \qquad \text{if } \mu \neq 0$$

and

$$\frac{1}{(k-1)!}\frac{\mathbb{P}_T[N_T^{(k)}]}{T^{k-1}}\frac{1}{T\mathbb{P}_T(N_T \geq k)} \to k\Big(\frac{r\sigma}{2}\Big)^k \qquad \text{if } \mu = 0.$$

We deduce that

$$\begin{aligned}\mathbb{P}_T&\bigg[\frac{1}{N_T^{(k)}}\sum_{u\in\mathcal{N}_T^{(k)}} f(u)\,\bigg|\, N_T \geq k\bigg]\\ &= (1+o(1))k\Big(\frac{\sigma}{2\mu}\Big)^k(e^{r\mu}-1)^k\int_0^\infty (T(1-e^{-\phi/T}))^{k-1}\mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k)/T}f(\xi_T)]\,\mathrm{d}\phi \qquad (3.29)\end{aligned}$$

when $\mu \neq 0$, and when $\mu = 0$

$$\begin{aligned}\mathbb{P}_T&\bigg[\frac{1}{N_T^{(k)}}\sum_{u\in\mathcal{N}_T^{(k)}} f(u)\,\bigg|\, N_T \geq k\bigg]\\ &= (1+o(1))k\Big(\frac{r\sigma}{2}\Big)^k\int_0^\infty (T(1-e^{-\phi/T}))^{k-1}\mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k)/T}f(\xi_T)]\,\mathrm{d}\phi.\end{aligned}$$

Our aim now is to choose f as in Lemma 3.6.12, and apply dominated convergence and Lemma 3.6.12 to complete the proof. We do this only in the case $\mu \neq 0$; the case $\mu = 0$ is very similar. Let

$$A(\phi,T) = (T(1-e^{-\phi/T}))^{k-1}\mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k)/T}f(\xi_T)]$$

and

$$B(\phi,T) = (T(1-e^{-\phi/T}))^{k-1}\mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k)/T}].$$

Then $0 \leq A(\phi,T) \leq B(\phi,T)$ for all ϕ, T. By letting $s_1,\ldots,s_{k-1} \downarrow 0$ in Lemma 3.6.12, we get that

$$\begin{aligned}\lim_{T\to\infty}\mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k)/T}\mathbb{1}_{\Upsilon_{k-1}}] &= \Big(\frac{1}{e^{r\mu}-1}\Big)^{k-1}\frac{1}{(1+\frac{\sigma}{2\mu}\phi(e^{r\mu}-1))^2}\Big(\frac{e^{r\mu}-1}{1+\frac{\sigma}{2\mu}\phi(e^{r\mu}-1)}\Big)^{k-1}\\ &= \frac{1}{(1+\frac{\sigma}{2\mu}\phi(e^{r\mu}-1))^{k+1}}.\end{aligned}$$

Also, by Lemma 3.6.6,

$$\mathbb{Q}_T^{k,T}[e^{-\phi(N_T-k)/T}\mathbb{1}_{\Upsilon_{k-1}^c}] \le \mathbb{Q}_T^{k,T}(\Upsilon_{k-1}^c) \to 0,$$

so

$$\lim_{T\to\infty} B(\phi,T) = \phi^{k-1}\frac{1}{(1+\frac{\sigma}{2\mu}\phi(e^{r\mu}-1))^{k+1}}.$$

On the other hand, by (3.29) with $f \equiv 1$,

$$1 = \mathbb{P}_T\left[\frac{1}{N_T^{(k)}}\sum_{u\in\mathcal{N}_T^{(k)}} 1 \,\middle|\, N_T \ge k\right] = (1+o(1))k\Big(\frac{\sigma}{2\mu}\Big)^k (e^{r\mu}-1)^k \int_0^\infty B(\phi,T)\,\mathrm{d}\phi,$$

so

$$\lim_{T\to\infty}\int_0^\infty B(\phi,T)\,\mathrm{d}\phi = \frac{1}{k}\Big(\frac{2\mu}{\sigma(e^{r\mu}-1)}\Big)^k;$$

and as a result we see that

$$\lim_{T\to\infty}\int_0^\infty B(\phi,T)\,\mathrm{d}\phi = \int_0^\infty \lim_{T\to\infty} B(\phi,T)\,\mathrm{d}\phi.$$

Therefore, by dominated convergence,

$$\lim_{T\to\infty}\int_0^\infty A(\phi,T)\,\mathrm{d}\phi = \int_0^\infty \lim_{T\to\infty} A(\phi,T)\,\mathrm{d}\phi. \tag{3.30}$$

Lemma 3.6.12 tells us that

$$A(\phi,T) \to \Big(\frac{\phi}{e^{r\mu}-1}\Big)^{k-1}\frac{h}{(1+\frac{\sigma}{2\mu}\phi(e^{r\mu}-1))^2}\prod_{i=1}^{k-1}\frac{e^{r\mu(1-s_i)}-1}{1+\frac{\sigma}{2\mu}\phi(e^{r\mu(1-s_i)}-1)}$$

where $h = \lim_{T\to\infty}\mathbb{Q}_T^{k,T}(H)$, so by (3.29) and (3.30),

$$\begin{aligned}
&\lim_{T\to\infty}\mathbb{P}_T\left[\frac{1}{N_T^{(k)}}\sum_{u\in\mathcal{N}_T^{(k)}} f(u) \,\middle|\, N_T\ge k\right]\\
&\quad= k\Big(\frac{\sigma}{2\mu}\Big)^k (e^{r\mu}-1)\int_0^\infty \phi^{k-1}\frac{h}{(1+\frac{\sigma}{2\mu}\phi(e^{r\mu}-1))^2}\prod_{i=1}^{k-1}\frac{e^{r\mu(1-s_i)}-1}{1+\frac{\sigma}{2\mu}\phi(e^{r\mu(1-s_i)}-1)}\,\mathrm{d}\phi\\
&\quad= \frac{k\sigma}{2\mu}(e^{r\mu}-1)\int_0^\infty \frac{h}{(1+\frac{\sigma}{2\mu}\phi(e^{r\mu}-1))^2}\prod_{i=1}^{k-1}\Big(1-\frac{1}{1+\frac{\sigma}{2\mu}\phi(e^{r\mu(1-s_i)}-1)}\Big)\,\mathrm{d}\phi.
\end{aligned}$$

Note that, for any $\mu \ne 0$, we have $\frac{\sigma}{2\mu}(e^{r\mu(1-s_i)}-1) > 0$ for all i, so we can apply the first part of Lemma 3.7.1 to get

$$\begin{aligned}
&\lim_{T\to\infty}\mathbb{P}_T\left[\frac{1}{N_T^{(k)}}\sum_{u\in\mathcal{N}_T^{(k)}} f(u) \,\middle|\, N_T\ge k\right]\\
&\qquad\qquad= hk\Big(\prod_{i=1}^{k-1}\frac{e_i}{e_i-e_0}\Big) + hke_0\sum_{j=1}^{k-1}\frac{e_j}{(e_j-e_0)^2}\Big(\prod_{\substack{i=1\\ i\ne j}}^{k-1}\frac{e_i}{e_i-e_j}\Big)\log\frac{e_0}{e_j}
\end{aligned}$$

where $e_j = \frac{\sigma}{2\mu}(e^{r\mu(1-s_j)}-1)$ for each j (including $j=0$, where $s_0=0$). □

3.7. Appendix

Here we gather some results that are easy but still require proofs. We begin with the calculation of some integrals.

Lemma 3.7.1. Suppose that $k \geq 2$ and $e_0, \ldots, e_{k-1} \in (0, \infty)$ with $e_i \neq e_j$ for any $i \neq j$. Then

$$\int_0^\infty \frac{1}{(1+\theta e_0)^2} \prod_{j=1}^{k-1} \Big(1 - \frac{1}{1+\theta e_j}\Big) \mathrm{d}\theta$$
$$= \frac{1}{e_0} \Big(\prod_{i=1}^{k-1} \frac{e_i}{e_i - e_0} \Big) + \sum_{j=1}^{k-1} \frac{e_j}{(e_j - e_0)^2} \Big(\prod_{\substack{i=1 \\ i \neq j}}^{k-1} \frac{e_i}{e_i - e_j} \Big) \log \Big(\frac{e_0}{e_j} \Big)$$

and

$$\int_0^1 \frac{1}{(1+\theta e_0)^2} \prod_{j=1}^{k-1} \Big(1 - \frac{1}{1+\theta e_j}\Big) \mathrm{d}\theta$$
$$= \frac{1}{1+e_0} \prod_{i=1}^{k-1} \frac{e_i}{e_i - e_0} + \sum_{j=1}^{k-1} \frac{e_j}{(e_j - e_0)^2} \Big(\prod_{\substack{i=1 \\ i \neq j}}^{k-1} \frac{e_i}{e_i - e_j} \Big) \log \Big(\frac{1+e_0}{1+e_j} \Big).$$

Proof. First note that since $e_j \in (0, \infty)$ for each j,

$$0 \leq \int_0^\infty \frac{1}{(1+\theta e_0)^2} \prod_{j=1}^{k-1} \Big(1 - \frac{1}{1+\theta e_j}\Big) \mathrm{d}\theta \leq \int_0^\infty \frac{1}{(1+\theta e_0)^2} \mathrm{d}\theta < \infty.$$

Expanding the product, we have

$$\prod_{j=1}^{k-1} \Big(1 - \frac{1}{1+\theta e_j}\Big) = 1 - \sum_{j_1=1}^{k-1} \frac{1}{1+\theta e_{j_1}} + \sum_{j_1 < j_2} \prod_{i=1}^{2} \frac{1}{1+\theta e_{j_i}} + \ldots + \sum_{j_1 < \ldots < j_{k-1}} \prod_{i=1}^{k-1} \frac{1}{1+\theta e_{j_i}}.$$

We view this as one sum in which all terms are products of factors of the form $\frac{1}{1+\theta e_i}$ for some i; therefore, using partial fractions, the whole thing can be written as a sum of terms of the form $\frac{c_i}{1+\theta e_i}$ for some coefficients c_i which do not depend on θ. As a result, our entire integrand may be written in the form

$$\frac{1}{(1+\theta e_0)^2} \prod_{j=1}^{k-1} \Big(1 - \frac{1}{1+\theta e_j}\Big) = \frac{a_1}{1+\theta e_0} + \frac{a_2}{(1+\theta e_0)^2} + \sum_{j=1}^{k-1} \frac{b_j}{1+\theta e_j} \tag{3.31}$$

for some coefficients a_1, a_2 and $b_1, \ldots, b_{k-1}$ that do not depend on θ.

Setting $\theta = -1/e_0$, we see that necessarily

$$a_2 = \prod_{i=1}^{k-1} \frac{e_i}{e_i - e_0}.$$

Setting $\theta = -1/e_j$ for $j = 1, \ldots, k-1$, some elementary calculations reveal that

$$b_j = -\frac{e_j^2}{(e_j - e_0)^2} \prod_{\substack{i=1 \\ i \neq j}}^{k-1} \frac{e_i}{e_i - e_j}.$$

For a_1, we observe that

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\prod_{j=1}^{k-1}\Big(1-\frac{1}{1+\theta e_j}\Big) \to a_1 e_0 \quad \text{as } \theta \to -1/e_0,$$

and then after some simple calculations we get

$$a_1 = e_0\Big(\prod_{i=1}^{k-1}\frac{e_i}{e_i-e_0}\Big)\sum_{j=1}^{k-1}\frac{1}{e_j-e_0}.$$

Integrating (3.31),

$$\begin{aligned}\int_0^\infty \frac{1}{(1+\theta e_0)^2}\prod_{j=1}^{k-1}\Big(1-\frac{1}{1+\theta e_j}\Big)\mathrm{d}\theta &= \lim_{M\to\infty}\Big(\frac{a_1}{e_0}\log(1+Me_0)+\frac{a_2}{e_0}+\sum_{j=1}^{k-1}\frac{b_j}{e_j}\log(1+Me_j)\Big)\\ &= \frac{a_1}{e_0}\log e_0+\frac{a_2}{e_0}+\sum_{j=1}^{k-1}\frac{b_j}{e_j}\log e_j+\lim_{M\to\infty}\Big(\frac{a_1}{e_0}+\sum_{j=1}^{k-1}\frac{b_j}{e_j}\Big)\log M.\end{aligned}$$

Since we have already checked that the integral is finite, the last term must be zero; this leaves

$$\begin{aligned}\int_0^\infty \frac{1}{(1+\theta e_0)^2}\prod_{j=1}^{k-1}\Big(1-\frac{1}{1+\theta e_j}\Big)\mathrm{d}\theta &= \frac{a_1}{e_0}\log e_0+\frac{a_2}{e_0}+\sum_{j=1}^{k-1}\frac{b_j}{e_j}\log e_j\\ &= -\sum_{j=1}^{k-1}\frac{b_j}{e_j}\log e_0+\frac{a_2}{e_0}+\sum_{j=1}^{k-1}\frac{b_j}{e_j}\log e_j\end{aligned}$$

which is the first part of the result. The second part follows similarly by integrating (3.31) over $(0,1)$ instead of $(0,\infty)$. □

Lemma 3.7.2. For any $0 \le s_j \le T$, $\beta \ne \alpha$ and $y \in [0,1]$,

$$\int_{s_j}^T \frac{e^{(\beta-\alpha)(T-s)}}{(\beta(1-y)e^{(\beta-\alpha)(T-s)}+\beta y-\alpha)^2}\mathrm{d}s = \frac{e^{(\beta-\alpha)(T-s_j)}-1}{(\beta-\alpha)^2(\beta(1-y)e^{(\beta-\alpha)(T-s_j)}+\beta y-\alpha)}.$$

Also, for any $0 \le s_i \le 1$, $r,\sigma > 0$ and $\mu \ne 0$,

$$\int_{s_i}^1 \frac{e^{r\mu(1-s)}}{(1+\frac{\sigma}{2\mu}\phi(e^{r\mu(1-s)}-1))^2}\mathrm{d}s = \frac{1}{r\mu}\Big(\frac{e^{r\mu(1-s_i)}-1}{1+\frac{\sigma}{2\mu}\phi(e^{r\mu(1-s_i)}-1)}\Big).$$

Proof. By substituting $t = e^{(\beta-\alpha)(T-s)}$, we see that

$$\begin{aligned}\int_{s_j}^T \frac{e^{(\beta-\alpha)(T-s)}}{(\beta(1-y)e^{(\beta-\alpha)(T-s)}+\beta y-\alpha)^2}\mathrm{d}s &= \frac{1}{\beta-\alpha}\int_1^{e^{(\beta-\alpha)(T-s_j)}}\frac{1}{(\beta(1-y)t+\beta y-\alpha)^2}\mathrm{d}t\\ &= \frac{1}{(\beta-\alpha)\beta(1-y)}\Big(\frac{1}{\beta-\alpha}-\frac{1}{\beta(1-y)e^{(\beta-\alpha)(T-s_j)}+\beta y-\alpha}\Big).\end{aligned}$$

Furthermore,

$$\begin{aligned}\frac{1}{\beta-\alpha}-\frac{1}{\beta(1-y)e^{(\beta-\alpha)(T-s_j)}+\beta y-\alpha} &= \frac{\beta(1-y)e^{(\beta-\alpha)(T-s_j)}+\beta y-\beta}{(\beta-\alpha)(\beta(1-y)e^{(\beta-\alpha)(T-s_j)}+\beta y-\alpha)}\\ &= \frac{\beta(1-y)(e^{(\beta-\alpha)(T-s_j)}-1)}{(\beta-\alpha)(\beta(1-y)e^{(\beta-\alpha)(T-s_j)}+\beta y-\alpha)}.\end{aligned}$$

Combining these two calculations gives the first part of the result. The second is very similar. □

The following lemma is elementary, but we do not know a suitable reference.

Lemma 3.7.3. Suppose that $X_1 \le X_2 \le \ldots \le X_n$ are ordered random variables satisfying

$$\mathbb{P}(X_1 \in (a_1, b_1], \ldots, X_n \in (a_n, b_n]) = \int_{a_1}^{b_1} \cdots \int_{a_1}^{b_1} f(x_1, \ldots, x_n) \mathrm{d}x_n \ldots \mathrm{d}x_1$$

for some symmetric function f and any $a_1 < b_1 \le a_2 < b_2 \le \ldots \le a_n < b_n$. Let $Y_1, \ldots, Y_n$ be a uniformly random permutation of $X_1, \ldots, X_n$. Then

$$\mathbb{P}(Y_1 > y_1, \ldots, Y_n > y_n,\, Y_i \neq Y_j \,\forall i \neq j) = \frac{1}{n!} \int_{y_1}^{1} \cdots \int_{y_n}^{1} f(x_1, \ldots, x_n) \mathrm{d}x_n \ldots \mathrm{d}x_1.$$

Proof. First note that, via a standard limiting procedure, for any $c_1, \ldots, c_n \in [0, 1]$,

$$\mathbb{P}(X_1 > c_1, \ldots, X_n > c_n,\, X_i \neq X_j \,\forall i \neq j) = \int_{c_1}^{1} \cdots \int_{c_n}^{1} f(x_1, \ldots, x_n) \mathbb{1}_{\{x_1 < \ldots < x_n\}} \mathrm{d}x_n \ldots \mathrm{d}x_1.$$

We now deviate from our usual notation by temporarily letting S_n be the symmetric group on n objects. Then

$$\begin{aligned}
&\mathbb{P}(Y_1 > y_1, \ldots, Y_n > y_n,\, Y_i \neq Y_j \,\forall i \neq j) \\
&\qquad = \sum_{\sigma \in S_n} \frac{1}{n!} \mathbb{P}(\sigma(X_1) > y_1, \ldots, \sigma(X_n) > y_n,\, X_i \neq X_j \,\forall i \neq j) \\
&\qquad = \frac{1}{n!} \sum_{\sigma \in S_n} \mathbb{P}(X_1 > \sigma^{-1}(y_1), \ldots, X_n > \sigma^{-1}(y_n),\, X_i \neq X_j \,\forall i \neq j) \\
&\qquad = \frac{1}{n!} \sum_{\sigma \in S_n} \int_{\sigma^{-1}(y_1)}^{1} \cdots \int_{\sigma^{-1}(y_n)}^{1} f(x_1, \ldots, x_n) \mathbb{1}_{\{x_1 < \ldots < x_n\}} \mathrm{d}x_n \ldots \mathrm{d}x_1 \\
&\qquad = \frac{1}{n!} \sum_{\sigma \in S_n} \int_{y_1}^{1} \cdots \int_{y_n}^{1} f(\sigma(x_1), \ldots, \sigma(x_n)) \mathbb{1}_{\{\sigma(x_1) < \ldots < \sigma(x_n)\}} \mathrm{d}x_n \ldots \mathrm{d}x_1.
\end{aligned}$$

Since f is symmetric, this equals

$$\frac{1}{n!} \sum_{\sigma \in S_n} \int_{y_1}^{1} \cdots \int_{y_n}^{1} f(x_1, \ldots, x_n) \mathbb{1}_{\{\sigma(x_1) < \ldots < \sigma(x_n)\}} \mathrm{d}x_n \ldots \mathrm{d}x_1,$$

and since for any $x_1, \ldots, x_n$, exactly one of the permutations in S_n satisfies $\sigma(x_1) < \ldots < \sigma(x_n)$, we get the result. ☐

Finally, we prove Lemma 3.6.1. This roughly said that we can assume without loss of generality that $\mathbb{P}_T[L^{(j)}] = o(T^{j-2})$ for each $j \ge 3$. More precisely, for each $k \ge 1$, under $\mathbb{P}_T$, there exists a coupling between our Galton-Watson tree with offspring distribution L (and its k chosen particles) and another Galton-Watson tree with offspring distribution $\tilde{L}$ satisfying

- $\mathbb{P}_T[\tilde{L}] = 1 + \mu/T + o(1/T)$;
- $\mathbb{P}_T[\tilde{L}(\tilde{L} - 1)] = \sigma + o(1)$;
- $\mathbb{P}_T[\tilde{L}^{(j)}] = o(T^{j-2})$ for all $j \ge 3$,

such that with probability tending to 1, the two trees induced by the k chosen particles are equal until time T.

Proof of Lemma 3.6.1. We claim that we can choose integers $J(T)$ such that $J(T) = o(T)$ and $\sum_{j=J(T)}^{\infty} jp_j^{(T)} = o(1/T)$. To see this, first note that for any $\varepsilon > 0$, $\sum_{\varepsilon T}^{\infty} jp_j^{(T)} = o(1/T)$, otherwise $\sum_{\varepsilon T}^{\infty} j^2 p_j^{(T)}$ is larger than a constant infinitely often, contradicting the uniform integrability of L^2.

Choose any sequence $\varepsilon_i \to 0$; by the above, we may choose $t_i \geq i$ such that

$$\sum_{\varepsilon_i T}^{\infty} jp_j^{(T)} < \frac{\varepsilon_i}{T} \quad \forall T \geq t_i. \tag{3.32}$$

Then for any T, let $I(T) = \max\{i : t_i \leq T\}$ and $J(T) = \lceil \varepsilon_{I(T)} T \rceil$.

Since $I(T) \to \infty$ (because $t_i \geq i$) we have $J(T) = \lceil \varepsilon_{I(T)} T \rceil = o(T)$. But also, by (3.32),

$$\sum_{j=J(T)}^{\infty} jp_j^{(T)} < \frac{\varepsilon_{I(T)}}{T}$$

since $T \geq t_{I(T)}$ by definition of $I(T)$. Therefore $J(T)$ satisfies the claim.

We now choose our distribution $\tilde{L}$. If $1 \leq j < J(T)$ then let $\tilde{p}_j^{(T)} = p_j^{(T)}$. If $j \geq J(T)$ then let $\tilde{p}_j^{(T)} = 0$. Then choose $\tilde{p}_0^{(T)}$ so that $\sum_j \tilde{p}_j^{(T)} = 1$. Let $\tilde{L}$ satisfy

$$\mathbb{P}_T(\tilde{L} = j) = \tilde{p}_j^{(T)} \quad \forall j \geq 0.$$

We then have

$$\mathbb{P}_T[\tilde{L}] = \sum_{j=1}^{J(T)-1} jp_j^{(T)} = \mathbb{P}_T[L] - \sum_{j=J(T)}^{\infty} jp_j^{(T)} = 1 + \frac{\mu}{T} + o(1/T)$$

by the claim that we have just proved about $J(T)$,

$$\mathbb{P}_T[\tilde{L}(\tilde{L}-1)] = \sum_{j=2}^{J(T)-1} j(j-1)p_j^{(T)} = \mathbb{P}_T[L(L-1)] - \sum_{j=J(T)}^{\infty} j(j-1)p_j^{(T)} = \sigma + o(1)$$

by the fact that L^2 is uniformly integrable, and

$$\mathbb{P}_T[\tilde{L}^{(i)}] = \sum_{j=i}^{J(T)-1} j^{(i)} p_j^{(T)} \leq J(T)^{i-2} \sum_{j=2}^{\infty} j(j-1)p_j^{(T)} = J(T)^{i-2}(\sigma + o(1)) = o(T^{i-2})$$

since $J(T) = o(T)$. Therefore $\tilde{L}$ satisfies the three properties required in the statement of the lemma.

Couple two Galton-Watson trees $\mathrm{GW}(L)$ and $\mathrm{GW}(\tilde{L})$ in the obvious way: if a particle in $\mathrm{GW}(L)$ has j children for some $j < J(T)$, then it also has j children in $\mathrm{GW}(\tilde{L})$. On the other hand, if a particle in $\mathrm{GW}(L)$ has j children for some $j \geq J(T)$, then it has no children in $\mathrm{GW}(\tilde{L})$. The set of particles in $\mathrm{GW}(\tilde{L})$ is then a subset of those in $\mathrm{GW}(L)$ and any particle that exists in $\mathrm{GW}(\tilde{L})$ has lifetime equal to its counterpart in $\mathrm{GW}(L)$. Choose k particles uniformly at random without replacement at time T in $\mathrm{GW}(L)$. If they exist in $\mathrm{GW}(\tilde{L})$ then they are also our chosen particles in $\mathrm{GW}(\tilde{L})$; if not, then pick k particles uniformly and independently from $\mathrm{GW}(\tilde{L})$.

The two trees induced by the chosen particles are equal if and only if none of the ancestors of the k chosen particles in $\mathrm{GW}(L)$ gave birth to more than $J(T)$ children. By a union bound, it suffices to show that the probability that the first of the k particles has an ancestor that gave birth to more than $J(T)$ particles, conditional on $N_T \geq k$, tends to 0.

For a particle $u \in \mathcal{N}_T$, let $\Phi_T(u)$ be the event that at least one of the ancestors of u had more than $J(T)$ children. By (3.6),

$$\mathbb{P}_T\Big[\frac{1}{N_T}\sum_{u\in\mathcal{N}_T}\mathbb{1}_{\Phi_T(u)}\,\Big|\,N_T\geq 1\Big] = \frac{\mathbb{P}_T[N_T]}{\mathbb{P}_T(N_T\geq 1)}\mathbb{Q}_T^{1,T}\Big[\frac{1}{N_T}\mathbb{1}_{\Phi_T(\xi_T^1)}\Big].$$

By the FKG inequality,

$$\mathbb{Q}_T^{1,T}\Big[\frac{1}{N_T}\mathbb{1}_{\Phi_T(\xi_T^1)}\Big] \leq \mathbb{Q}_T^{1,T}\Big[\frac{1}{N_T}\Big]\mathbb{Q}_T^{1,T}(\Phi_T(\xi_T^1)),$$

so applying (3.6) again with $f\equiv 1$,

$$\mathbb{P}_T\Big[\frac{1}{N_T}\sum_{u\in\mathcal{N}_T}\mathbb{1}_{\Phi_T(u)}\,\Big|\,N_T\geq 1\Big] \leq \frac{\mathbb{P}_T[N_T]}{\mathbb{P}_T(N_T\geq 1)}\mathbb{Q}_T^{1,T}\Big[\frac{1}{N_T}\Big]\mathbb{Q}_T^{1,T}(\Phi_T(\xi_T^1)) = \mathbb{Q}_T^{1,T}(\Phi_T(\xi_T^1)).$$

By Markov's inequality, this is at most the expected number of births of size larger than $J(T)$ along the spine by time T under $\mathbb{Q}_T^{1,T}$; by Lemma 3.4.7 (note that since we have only one spine, $\psi_1=\infty$) the births occur as a Poisson point process of rate rm_T, and by Lemma 3.4.8 the sizes of the births are size-biased. Thus

$$\mathbb{Q}_T^{1,T}(\Phi_T(\xi_T^1)) \leq rm_T T\sum_{j=J(T)}^{\infty}\frac{jp_j^{(T)}}{m_T}.$$

But $m_T\to 1$ and we chose $J(T)$ such that the sum above is $o(1/T)$; so $\mathbb{Q}_T^{1,T}(\Phi_T(\xi_T^1))\to 0$ and therefore

$$\mathbb{P}_T\Big[\frac{1}{N_T}\sum_{u\in\mathcal{N}_T}\mathbb{1}_{\Phi_T(u)}\,\Big|\,N_T\geq 1\Big]\to 0.$$

We wanted to show that the probability that the first chosen particle has an ancestor that gave birth to more than $J(T)$ particles, conditional on $N_T\geq k$, tends to 0. We have shown the same statement conditional on $N_T\geq 1$, but a standard Markov chains argument shows that $\mathbb{P}_T(N_T\in\{1,\dots,k-1\})\to 0$ and the result follows. $\square$

Chapter 4

Heavy-tailed trees

4.1. Universality for heavy-tailed trees

We begin this section by looking at the coalescent structure of a class of supercritical trees. For $\alpha \in (0,1]$, consider the supercritical offspring variable $L^*_{1+\alpha}$ given by generating function

$$\mathbb{E}[s^{L_{1+\alpha}}] = f_{1+\alpha}(s) = \frac{(1-s)^{1+\alpha} - 1 + (1+\alpha)s}{\alpha}. \tag{4.1}$$

This random variable has a special connection with heavy-tailed critical processes which we explore shortly, but it is worth mentioning here it also has cameo appearences in [12] and in [8] in the analysis of Levy processes and beta coalescents respectively.

In the boundary case $\alpha = 0$, we just write $L^* = L^*_1$. This variable is characterised by generating function,

$$\mathbb{E}[s^{L_*}] = f_*(s) = s - (1-s)\log\left(\frac{1}{1-s}\right)$$

or more concretely

$$\mathbb{P}(L_* = n) = \frac{1}{n(n-1)},\ n = 2, 3, ...$$

Remarkably, we see below that L^* is the offspring distribution with the most harmonious coalescent structure. The following two theorems give the limiting coalescent structure of the $L^*_{1+\alpha}$ and L^* trees.

Theorem 4.1.1.

$$(\pi_t^{k,L_{1+\alpha},T})_{t\in[0,T]} \to^D (\bar{\pi}_t^{k,L_{1+\alpha}})_{t\in[0,\infty)}, \tag{4.2}$$

and $(\bar{\pi}_t^{k,L_{1+\alpha}})_{t\in[0,\infty)}$ has law given by

$$\mathbb{P}(\bar{\pi}^{k,L} : \eta_0 \prec^{dt_1} \eta_1 \prec^{dt_2} ... \prec^{dt_n} \eta_n) = \int_0^\infty \frac{\theta^{\alpha n}}{(k-1)!(1+\theta^\alpha)^{1+1/\alpha}} \prod_{i=1}^n \left\{ \frac{(1+\alpha)^{(c)}}{\alpha} \frac{e^{-t_i}}{(1+\theta^\alpha e^{-t_i})^c} dt_i \right\} dv. \tag{4.3}$$

Theorem 4.1.2.

$$(\rho_t^{k,L_*,T})_{t\in[0,T]} \to^D (\bar{\rho}^{k,\mathsf{BS}})_{t\in[0,\infty)}$$

Where $(\rho_t^{k,\mathsf{BS}})_{t\in[0,\infty)}$ is the Bolthausen-Sznitman coalescent restricted to $\{1,...,k\}$.

Though interesting results in their own right, the two theorems above are motivated by intimate connections with heavy-tailed trees. Our first theorem below states that the trees with offspring variables $\{L_{1+\alpha}, \alpha \in (0,1]\}$ have limiting coalescent structures that are simple time changes of universal coalescent limits for a large class of heavy-tailed critical trees. Our second result below states that we also that a class of extremely heavy-tailed critical trees have coalescents converging to various time-changes of the Bolthausen-Sznitman coalescent. These theorems are inspired by the incredible results of Lageras and Sagitov [30] which we discuss below after a few definitions.

Let us reintroduce in more details some concepts from the introduction. Let

$$\mathcal{R} = \left\{ \Delta : [1,\infty) \to \mathbb{R} \;\middle|\; \forall \lambda > 0 \lim_{x\to\infty} \frac{\Delta(\lambda x)}{\Delta(x)} = 1 \right\}$$

be the set of slowly varying functions on $[1,\infty)$. For $\alpha \in (0,1]$, let

$$\mathcal{M}^{1+\alpha} = \left\{ L \text{ critical offspring distribution} : \exists \Delta \in \mathcal{R} : \mathbb{E}[s^L] = s + (1-s)^{1+\alpha}\Delta\Big(\frac{1}{1-s}\Big) \right\}$$

be the set of offspring distributions in the domain of attraction of a $(1+\alpha)$-stable law. In the $\alpha = 1$ case we insist further that $\lim_{x\to\infty} \Delta(x) = x_0$ exists, and from here it is straightforward to verify from here that $\mathcal{M}^2$ is identical the collection of all critical offspring distributions with finite variance.

We are now equipped to state a first result from Lageras and Sagitov [30].

Theorem 4.1.3 (Lageras and Sagitov)**.** Fix T, and run a process $(N_t)_{t\in[0,T]}$ with offspring distribution $L \in \mathcal{M}^{1+\alpha}$ until time T. On the event $\{N_T > 0\}$, colour any particle that has a descendent alive at time T purple. Conditioned on the event that $N_T > 0$, let $(\tilde{N}_t)_{t\in[0,T]}$ be the process associated with the number of purple particles alive. Then as $T \to \infty$, the purple process can be described dynamically by the following convergence in Skorokhod topology

$$\left(\tilde{N}_{tT}\right)_{t\in[0,1]} \to^D \left(N^{L^*_{1+\alpha}}_{\log(\frac{1}{1-t})}\right)_{t\in[0,1]}, \tag{4.4}$$

where $\left(N_r^{L^*_{1+\alpha}}\right)_{r\in[0,\infty)}$ is a process with offspring distribution $L^*_{1+\alpha}$.

Since the coalescent structure of a tree depends precisely on its purple tree, this time change suggests heuristically that large-T coalescent properties of $\mathcal{M}^{1+\alpha}$-trees can be obtained by examining properties of the $L_{1+\alpha}$-trees, in that for $L \in \mathcal{M}^{1+\alpha}$

$$\left(\pi_{tT}^{k,L,T}\right)_{t\in[0,1]} \to^D \left(\bar{\pi}^{k,L^*_{1+\alpha}}_{\log(\frac{1}{1-t})}\right)_{t\in[0,1]} \tag{4.5}$$

The complete proof that the convergence (4.4) implies (4.5) is ongoing work. Working on the basis that (4.5) holds, combining this with Theorem 4.1.1 we have the following.

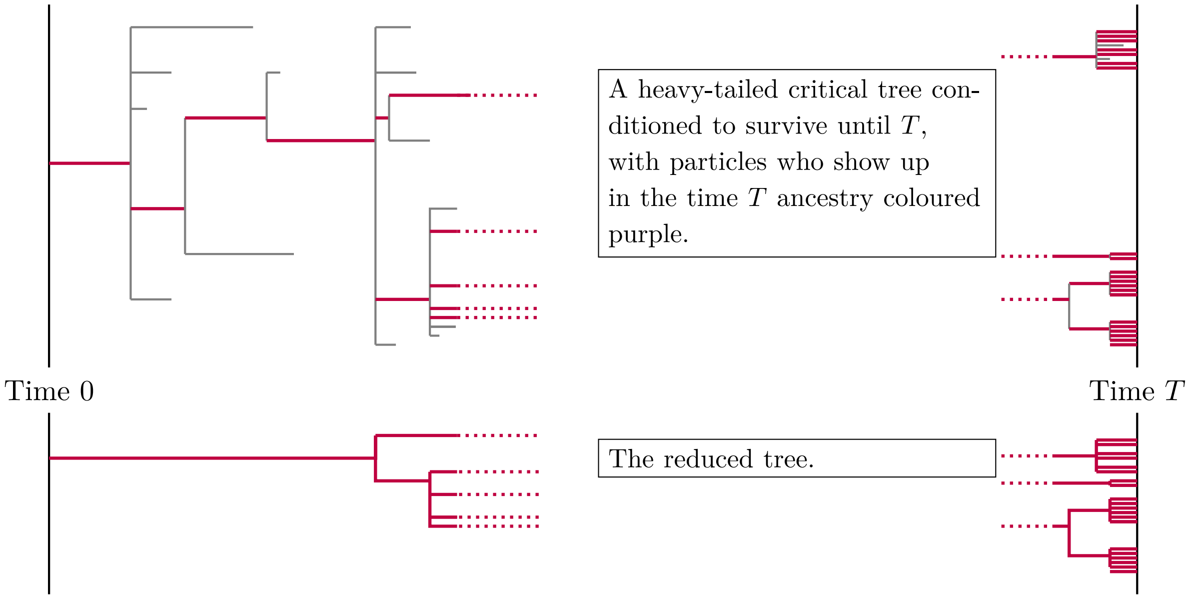


Figure 4-1: *Coalescent properties of heavy-tailed critical trees conditioned to survive can be understood through their reduced trees, which converge to time-changes of supercritical trees*

Theorem 4.1.4. The critical trees with offspring distributions in $\mathcal{M}^{1+\alpha}$ form a universal class with the same limiting coalescent structure, in that there exists a limit process $(\pi_t^{k,\alpha-\mathsf{Crit}})_{t\in[0,1]}$ such that for every $L \in \mathcal{M}^{1+\alpha}$,

$$(\pi_{tT}^{k,L,T})_{t\in[0,1]} \to^D (\bar{\pi}_t^{k,\alpha-\mathsf{Crit}})_{t\in[0,1]}. \tag{4.6}$$

where $(\bar{\pi}_t^{k,\alpha-\mathsf{Crit}})_{t\in[0,1]}$ is a partition process with splitting times given by

$$\mathbb{P}(\bar{\pi}^{k,L} : \eta_0 \prec^{dt_1} \eta_1 \prec^{dt_2} \ldots \prec^{dt_n} \eta_n) = \int_0^\infty \frac{\theta^{\alpha n}}{(k-1)!(1+\theta^\alpha)^{1+1/\alpha}} \prod_{i=1}^n \left\{ \frac{(1+\alpha)^{(c)}}{\alpha} \frac{1}{(1+\theta^\alpha(1-t_i))^c} dt_i \right\} dv. \tag{4.7}$$

Let us now consider what happens when $\alpha = 0$, where we require a finer analysis of the tail asymptotics. Indeed, for $\beta > 0$, let $\mathcal{M}_\beta^1$ be the set of offspring generating functions of the form

$$f(s) = s + (1-s)\Delta\left(\frac{1}{1-s}\right)$$

where Δ satisfies Zubkov's regularity condition

$$\Delta(x) \sim \log(x)^{-\beta}\Delta_1(\log(x))$$

for some other slowly varying Δ_1. We now state a second time-change result from [30].

Theorem 4.1.5 (Lageras and Sagitov)**.** Let $L \in \mathcal{M}_\beta^1$, then if $(N_t)_{t\in[0,T]}$ is a process with

$$(N_{tT})_{t\in[0,1]} \to^D \left(N^{L_*}_{\frac{\beta}{1+\beta}\log\left(\frac{1}{1-t}\right)}\right)_{t\in[0,1]}$$

where $(N_t^{L^*})_{t\in[0,\infty)}$ is a process with offspring distribution L^*, and the convergence holds in the same sense it did above.

Again, in light of this time-change, we anticipate that for any $L \in \mathcal{M}_\beta^1$,

$$\left(\rho_{tT}^{k,L,T}\right)_{t\in[0,1]} \to^D \left(\bar{\rho}^{k,L_*}_{\frac{\beta}{1+\beta}\log\left(\frac{1}{1-t}\right)}\right)_{t\in[0,1]}$$

and since the L^*-coalescent is Bolthausen-Sznitmann in limit, we arrive at the following.

Theorem 4.1.6. For each $\beta > 0$, the $\mathcal{M}_\beta^1$-trees form a universal class whose coalescents converge to a β-dependent time-change of the Bolthausen-Sznitman coalescent. That is, for $L \in \mathcal{M}_\beta^1$,

$$\left(\rho_{tT}^{k,L,T}\right)_{t\in[0,1]} \to^D \left(\rho^{k,\mathsf{BS}}_{\frac{\beta}{\beta+1}\log\left(\frac{1}{1-t}\right)}\right)_{t\in[0,1]}.$$

Section 4.2 is dedicated to the $\alpha \in (0,1]$ case. Section 4.3 looks at $\alpha = 0$.

4.2. Critical trees with infinite variance

Proof of Theorem 1.1.1. Consider the generating function

$$f_{1+\alpha}(s) = \frac{(1-s)^{1+\alpha} - 1 + (1+\alpha)s}{\alpha}.$$

By the martingale limit equation (**??**), a calculation shows that the Laplace transform of the $L_{1+\alpha}$ martingale limit has the form

$$\varphi_{1+\alpha}(\theta) = 1 - (1+\theta^{-\alpha})^{-1/\alpha}$$

Finally, for $p \geq 2$, $f_{1+\alpha}^p(s) = \frac{(\alpha+1)^{(p)}}{\alpha}\frac{1}{(1-s)^{p-(1+\alpha)}}$. The result follows after plugging these parts into the supercritical theorem of [24]. □

Now (4.7) is obtained from (4.3) simply by using the time change

$$u(t) = \log\left(\frac{1}{1-t}\right)$$

to make the substitution

$$\frac{e^{-u_i}}{(1+\theta^\alpha e^{-u_i})^c}du_i \to \frac{1}{(1+\theta^\alpha(1-t_i))^c}dt_i.$$

4.3. Infinite mean trees

In this section we will go so far as to calculate the finite dimensional distributions of the process $(\rho_t^{k,L^*,T})_{t\in[0,T]}$ for fixed times T. We then show these finite dimensional distributions converge as $T \to \infty$, and that the limit process is Bolthausen-Sznitman by examining its infinitesimal jump rates. First, we look at the process generating function.

Lemma 4.3.1. Let $F_t(s) = \mathbb{E}[s^{N_t}]$. Then

$$F_t(s) = 1 - (1-s)^{e^{-t}}$$

Proof. $F_t(s)$ satisfies the partial differential equation

$$\frac{\partial}{\partial t} F_t(s) = f(F) - F, \;\; F_0(s) = s,$$

see [5, Chapter III, Section 3] for details. $\square$

Observe here that $F_t(s)$, while ungainly as a function of t, has s-derivatives

$$F_u^k(s) = \frac{e^{-u}\prod_{i=1}^{k-1}(i - e^{-u})}{(1-s)^{(k-e^{-u})}}, \;\; u \geq 0. \tag{4.8}$$

Even more promising is the closed-form composition structure

$$F_u^k(F_v(s)) = \frac{e^{-u}\prod_{i=1}^{k-1}(i - e^{-u})}{(1-s)^{e^{-v}(k-e^{-u})}}, \;\; u, v \geq 0, \tag{4.9}$$

allowing for tractable insertion into our finite dimensional equations. Before we calculate the finite dimensional distributions, we need a quick combinatorial lemma.

Lemma 4.3.2. Let $0 = t_0 < t_1 < ... < t_n < t_{n+1} = T$, and let $[\{1, ..., k\}] = \gamma_0 \prec \gamma_1 \prec ... \prec \gamma_{n+1} = [\{1\}, ..., \{k\}]$ be a partition chain with breakage numbers b_{ij}, then

$$\sum_{i=0}^{n} \sum_{\Gamma_{ij} \in \gamma_i} e^{-(T-t_{i+1})}(b_{ij} - e^{-\Delta t_i}) = k - e^{-T}$$

Proof. Note that $\sum_{\Gamma_{ij}\in\gamma_i} b_{ij} = |\gamma_{i+1}|$, hence

$$\begin{aligned} \sum_{i=0}^{n} \sum_{\Gamma_{ij} \in \gamma_i} e^{-(T-t_{i+1})}(b_{ij} - e^{-\Delta t_i}) &= \sum_{i=0}^{n} \left\{ |\gamma_{i+1}| e^{-(T-t_{i+1})} - |\gamma_i| e^{-(T-t_i)} \right\} \\ &= |\gamma_{n+1}| e^{-(T-t_{n+1})} - |\gamma_0| e^{-(T-t_0)} = k - e^{-T} \end{aligned}$$

$\square$

Theorem 4.3.3.

$$\mathbb{P}(\rho_{t_i}^{k,L_*,T} = \gamma_i \;\forall\; i) = \frac{e^{t_n}\prod_{l=1}^{|\gamma_n|-1}(l - e^{-(T-t_n)})\prod_{i=0}^{n-1}\prod_{\Gamma_{ij}\in\gamma_i} e^{-\Delta t_i}\prod_{l=1}^{b_{ij}-1}(l - e^{-\Delta u_i})}{(k-1)!\mathbb{P}(N_T \geq k)} \tag{4.10}$$

Proof. We calculate the finite dimensional distributions of $\pi^{k,L^*,T}$, from which (4.10) can be obtained by replacing the t_i with $T - t_i$ and γ_i with γ_{n+1-i}. By Theorem 2.2 of [24],

$$\mathbb{P}(\pi_{t_1} = \gamma_1, ..., \pi_{t_n} = \gamma_n) = \int_0^1 \frac{(1-s)^{k-1}}{(k-1)!\mathbb{P}(N_T \geq k)} \prod_{i=0}^{i=n} \prod_{\Gamma_{i,j}\in\gamma_i} F_{\Delta t_i}^{b_{i,j}}\Big(F_{T-t_{i+1}}(s)\Big) ds. \tag{4.11}$$

Plugging in (4.9), we obtain

$$= \frac{\prod_{i=0}^{n}\prod_{\Gamma_{ij}\in\gamma_i}\left\{e^{-\Delta t_i}\prod_{l=1}^{l=b_{ij}-1}(l - e^{-\Delta t_i})\right\}}{(k-1)!\mathbb{P}(N_T \geq k)} \int_0^1 (1-s)^{\left\{k-1-\sum_{i=0}^{n}\sum_{\Gamma_{ij}\in\gamma_i} e^{-(T-t_{i+1})}(b_{ij} - e^{-\Delta t_i})\right\}} ds$$

Now apply lemma 4.3.2, and note that $\int_0^1 (1-s)^{-1+e^{-T}} ds = e^T$. $\square$

Taking $T \to \infty$ in (4.10), it is clear that we have the distributional convergence

$$\left(\rho_t^{k,L^*,T}\right)_{t\in[0,T]} \to^D \left(\bar{\rho}^{k,L^*}\right)_{t\in[0,\infty)}$$

where the limit process $\left(\bar{\rho}^{k,L^*}\right)_{t\in[0,\infty)}$ has finite dimensional distributions given by

$$\mathbb{P}(\bar{\rho}_{t_i}^{k,L_*} = \gamma_i \ \forall\ i) = \frac{(|\gamma_n|-1)!}{(k-1)!} e^{t_n} \prod_{i=0}^{n-1} \prod_{\Gamma_{ij}\in\gamma_i} \left\{ e^{-\Delta t_i} \prod_{l=1}^{b_{ij}-1} (l - e^{-\Delta t_i}) \right\} \tag{4.12}$$

What's now left to show is that (4.12) actually *are* the finite dimensional distributions of the Bolthausen-Sznitman coalescent. To this end, we need only show the limit process has the correct jump rates.

Lemma 4.3.4. Let γ be a partition of $\{1,2,...,k\}$ into p blocks, and let $\hat{\gamma} \prec \gamma$ be a partition obtained by merging m of the blocks of γ to form one block in $\hat{\gamma}$. Then

$$\lim_{h\to 0} \frac{1}{h} \mathbb{P}(\bar{\rho}_{t+h}^{k,L^*} = \hat{\gamma} | \bar{\rho}_t^{k,L^*} = \gamma) = \lambda_{p,m}$$

where

$$\lambda_{p,m} := \int_0^1 x^{m-2}(1-x)^{p-m} dx = \frac{(p-m)!(m-2)!}{(p-1)!},$$

Proof. The case $n = 1$ of (4.12) yields the one-dimensional distributions

$$\mathbb{P}(\bar{\rho}_t^{k,L^*} = \gamma) = \frac{(p-1)!}{(k-1)!} e^{-(p-1)t} \prod_{j=1}^{p} \left\{ \prod_{i=1}^{k_j} (i - e^{-t}) \right\}. \tag{4.13}$$

On the other hand, the two dimensional distributions give

$$\mathbb{P}(\bar{\rho}_t^{k,L^*} = \gamma, \bar{\rho}_{t+h}^{k,L^*} = \hat{\gamma}) = \frac{(p-m)!}{(k-1)!} e^{-(p-1)t} \prod_{j=1}^{p} \left\{ \prod_{i=1}^{k_j} (i - e^{-t}) \right\} e^{-(p-m)h} \prod_{i=1}^{m-1} (l - e^{-h}) \tag{4.14}$$

Dividing (4.14) by (4.13), we obtain

$$\mathbb{P}(\bar{\rho}_{t+h}^{k,L^*} = \hat{\gamma} | \bar{\rho}_t^{k,L^*} = \gamma) = \frac{(p-m)!}{(p-1)!} e^{-(p-m)h} \prod_{i=1}^{m-1} (l - e^{-h}),$$

differentiating with respect to h and sending $h \downarrow 0$ seals the result. $\square$

Bibliography

[1] David Aldous. The continuum random tree. I. *The Annals of Probability*, 19(1):1–28, 1991.

[2] David Aldous and Lea Popovic. A critical branching process model for biodiversity. *Advances in Applied Probability*, 37(04):1094–1115, 2005.

[3] K.B. Athreya. Coalescence in critical and subcritical Galton-Watson branching processes. *Journal of Applied Probability*, 49(3):627–638, 2012.

[4] K.B. Athreya. Coalescence in the recent past in rapidly growing populations *Stochastic Processes and their Applications*, 122 (2012) 3757–3766

[5] K.B. Athreya and P.E. Ney. *Branching Processes.* Springer-Verlag, New York, 1972.

[6] E. Bolthausen and A.-S. Sznitman On Ruelle's probability cascades and an abstract cavity method. Comm. Math. Phys. 197, 247-276. (1998)

[7] J. Berestycki, N. Berestycki and J. Schweinsberg. The genealogy of branching Brownian motion with absorption. *The Annals of Probability*, 2013, Vol. 41, No. 2, 527–618.

[8] J. Berestycki, N. Berestycki and J. Schweinsberg. Beta-coalescents and continuous stable random trees *The Annals of Probability*, 2007, Vol. 35, No. 5, 1835–1887.

[9] J. Bertoin and J-F. Le Gall The Bolthausen–Sznitman coalescent and the genealogy of continuous-state branching processes *Probability theory and related fields*, 117, 249–266 (2000).

[10] E. Bolthausen and A.-S. Sznitman. On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.* 197, 247-276.

[11] W. Buhler. Generations and degree of relationship in supercritical Markov branching processes. *Z. fur Wahr. und verw. Geb.* 18 (1971) 141-152.

[12] T. Dusquesne and J-F. Le Gall Random trees, Lévy processes and spatial branching processes. https://arxiv.org/pdf/math/0509558.pdf

[13] R Durrett. The genealogy of critical branching processes. *Stochastic Processes and their Applications*, 8(1):101–116, 1978.

[14] S.N. Ethier and T.G. Kurtz. *Markov processes: characterisation and convergence.* Wiley, 1986.

[15] K. Fleischmann and R. Siegmund-Schultze. The structure of reduced critical Galton-Watson processes *Math. Nachr.* 79, 233-241, 1977.

[16] Tanja Gernhard. The conditioned reconstructed process. *Journal of Theoretical Biology*, 253(4):769–778, 2008.

[17] N. Grosjean and T. Huillet On the genealogy and coalescence times of Bienayme-Galton-Watson branching processes. https://arxiv.org/abs/1709.07630

[18] T.E. Harris. *The Theory of Branching Processes*. Berlin, Springer. 1963.

[19] T.E. Harris. Some mathematical models of branching processes. The RAND corporation, 1951

[20] S.C. Harris, S.G.G. Johnston and M.I. Roberts. The coalescent structure of continuous-time Galton-Watson trees. https://arxiv.org/pdf/1703.00299.pdf

[21] S. C. Harris and M. I. Roberts. The many-to-few lemma and multiple spines. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(1):226–242, 2017.

[22] Simon C. Harris, Marion Hesse, and Andreas E. Kyprianou. Branching Brownian motion in a strip: survival near criticality. *The Annals of Probability*, 44(1):235–275, 2016.

[23] W.P. Johnson. The curious history of Faà di Bruno's formula. *The American Mathematical Monthly*, 109(3):217-234 · March 2002

[24] S.G.G. Johnston Coalescence in supercritical and subcritical continuous-time Galton-Watson trees. https://arxiv.org/abs/1709.08500

[25] O. Kallenberg. *Foundations of Modern Probability.* Springer-Verlag, New York, 1997

[26] S. Karlin and J. McGregor. Embeddability of discrete time simple branching processes into continuous time branching processes. TAMS 132, 115-136 (1968).

[27] H. Kesten and B. P. Stigum. A limit theorem for multidimensional Galton-Watson processes. AMS 37, 1211-1223

[28] J. F. C. Kingman. The coalescent. *Stochastic Processes and their Applications*. Volume 13, Issue 3, September 1982, Pages 235-248

[29] Andreas E. Kyprianou *Fluctuations of Lévy Processes with Applications* Springer-Verlag Berlin, Heidelberg, 2006.

[30] A. N. Lageras and S. Sagitov Reduced branching processes with very heavy tails *Journal of Applied Probability*, Vol. 45, No. 1 (Mar., 2008), pp. 190-200.

[31] Amaury Lambert. Coalescence times for the branching process. *Advances in Applied Probability*, 35(04):1071–1089, 2003.

[32] Amaury Lambert. The contour of splitting trees is a Lévy process. *The Annals of Probability*, 38(1):348–395, 2010.

[33] Amaury Lambert and Lea Popovic. The coalescent point process of branching trees. *The Annals of Applied Probability*, 23(1):99–144, 2013.

[34] Amaury Lambert. Probabilistic Models for the Subtrees of Life. https://arxiv.org/abs/1603.03705

[35] Amaury Lambert and Tanja Stadler. Birth–death models and coalescent point processes: The shape and probability of reconstructed phylogenies. *Theoretical Population Biology*, 90:113–128, 2013.

[36] Vi Le. Coalescence times for the Bienaymé-Galton-Watson process. *Journal of Applied Probability*, 51(01):209–218, 2014.

[37] J. Neveu A continuous-state branching process in relation with the GREM model of spin glass theory. *Res. Rep. 267, Ecole Polytechnique* (1992).

[38] Neil O'Connell. Yule process approximation of the skeleton of a branching process. *Journal of Applied Probability*, 41, 725–729, 1993.

[39] Neil O'Connell. The genealogy of branching processes and the age of our most recent common ancestor. *Advances in Applied Probability*, 27(02):418–442, 1995.

[40] A.G. Pakes Critical Markov branching process limit theorems allowing infinite variance. *Advances in Appied Probability*, Volume 42, Number 2 (2010), 460-488.

[41] É. Pardoux Continuous branching processes: the discrete hidden in the continuous. 2007 International Conference in Honor of Claude Lobry

[42] J. Pitman Coalescents with multiple collisions *The Annals of Probability*, 27, 1870-1902, 1999.

[43] Lea Popovic. Asymptotic genealogy of a critical branching process. *Annals of Applied Probability*, pages 2120–2148, 2004.

[44] L. C. G. Rogers. A guided tour through excursions. *Bull. London Math. Soc.*, 21:305–341, 1989.

[45] S. Sagitov and M. C.Serra Skeletons of near-critical Bienayme-Galton-Watson branching processes *Applied Probability Trust*, 2013.

[46] Jason Schweinsberg. Coalescent processes obtained from supercritical Galton–Watson processes. *Stochastic Processes and their Applications*, 106(1):107–139, 2003.

[47] E. Seneta. *Regularly Varying Functions.* Springer-Verlag. Berlin-Heidelberg-New York, 1976.

[48] A. L. Yakymiv Reduced branching processes *Theory Probab. Appl.* 25, 584-588. 1980

[49] A. M. Zubkov. Limiting Distributions of the Distance to the Closest Common Ancestor. *Theory Probab. Appl.*, 20(3), 602–612, 1974.